

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Пятигорский государственный университет»

Философские проблемы информационных технологий и киберпространства

Сетевой научный журнал

№ 2 (26) 2024

ISSN 2305-3763

<https://cyberspace.pgu.ru>

ФИЛОСОФСКИЕ ПРОБЛЕМЫ информационных технологий и киберпространства

№ 2 (26) 2024

ISSN 2305-3763

Свидетельство о регистрации
средства массовой информации
Эл. № ФС77-50786

<https://cyberspace.pgu.ru>

Сетевой журнал «Философские проблемы информационных технологий и киберпространства» является электронным научным изданием, официально зарегистрированным в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций (Свидетельство о регистрации средств массовой информации Эл № ФС77-50786).

Журнал руководствуется политикой свободного доступа (Open Access) на основании Лицензии Creative Commons «Attribution-NoDerivs» («Атрибуция – Без производных произведений») CC BY-ND.

Учредитель журнала – ФГБОУ ВО «Пятигорский государственный университет».

Издание включено в Российский индекс научного цитирования (РИНЦ) и представлено в Научной Электронной Библиотеке в открытом доступе, на открытой платформе научной электронной библиотеки Cyberleninka.ru и электронной библиотечной системе IPRBooks.

Журнал индексируется в международных базах данных: Ulrich's Periodicals Directory European Reference Index for the Humanities and the Social Sciences (ERIH PLUS), EBSCO Host, CrossRef (DOI), Social Science Open Access Repository (SSOAR) UlrichsWeb, EBSCOhost, а также в репозиториях CrossRef. Опубликованным статьям присваивается уникальный идентификатор DOI.

Первые издания увидели свет в 2010 г. в качестве сборников научных статей, выпущенных по итогам Международной конференции «Философские проблемы информационных технологий и киберпространства», регулярно проводимой на базе ФГБОУ ВО «Пятигорский государственный университет».

© ФГБОУ ВО «ПГУ», 2024

© Коллектив авторов, 2024

В статусе научного журнала издается с 2012 г.

Михайлов И.Ф.

Концепция рекурсии в когнитивных исследованиях.
Часть II: От Тьюринга к Байесу и к сознанию4

Хомяков А.

Аналогия как базовая функция мышления23

Стешук Б.И.

Трансформация политического участия в контексте появления
цифровых политических институтов42

Беседина А.Ю.

Эволюция методов обработки естественного языка52

Каспарян К.В., Рутковская М.В., Колесников И.Н.

Специфика отображения социальных сетей и компьютерных
технологий в фантастической вселенной Стивена Кинга
в начале 70-х гг. XX в. — начале 20-х гг. XXI в. в исторической
ретроспективе64

Хомяков А.Б.

Почему трансформеры не мыслят, как люди?87

УДК 004.8, 510.2, 519.2

DOI 10.17726/philIT.2024.2.1



Концепция рекурсии в когнитивных исследованиях. Часть II: От Тьюринга к Байесу и к сознанию

Михайлов Игорь Феликсович,

*доктор философских наук,
ведущий научный сотрудник,
Институт философии РАН
Москва, Россия*

ifmikhailov@iph.ras.ru

Аннотация. В статье обсуждается концепция рекурсии в математике, ИИ, когнитивных исследованиях и ее связь с сознанием. Развитие понятия прослеживается параллельно с историей теории вычислимости, когда были введены концепции машины Тьюринга с оракулом и вероятностной машины Тьюринга. Также рассматриваются такие рекурсивные вычислительные методы, как байесовская рекурсивная оценка и байесовский иерархический вывод. Показано, что с каждым нововведением в рекурсивных методах пределы вычислимости расширились. Автор утверждает, что рекурсия является жизненно важным аспектом человеческого познания, особенно в разработке и интерпретации сложного языка. В статье также характеризуются проблемы изучения рекурсии и сознания, такие как субъективная природа сознания и сложность нейронных сетей, связанных с сознательным мышлением. Кроме того, анализируются ограничения в понимании рекурсии современных теорий когнитивной обработки, освоения языка и сознания. Делается вывод о том, что исследование связи между рекурсией и сознанием имеет решающее значение для развития более глубокого понимания языка и когнитивной обработки. Автор ожидает, что будущая теория, основанная на рекурсии, поможет решить основные метафизические загадки прошлого и настоящего.

Ключевые слова: рекурсия; когнитивная обработка; сознание; усвоение языка; нейронные сети.

The Concept of Recursion in Cognitive Studies. Part II: From Turing to Bayes to Consciousness

Mikhailov Igor Felixovich,

*Doctor of Philosophy, Leading researcher,
Institute of Philosophy of the Russian Academy of Sciences
Moscow, Russian Federation*

ifmikhailov@iph.ras.ru

Abstract. This article discusses the concept of recursion in mathematics, AI, cognitive studies and its relationship to consciousness. The development of the notion is followed in parallel with the history of computability theory when concepts of Turing oracle and probabilistic machines were introduced. Also, such recursive computational techniques as Bayesian Recursive Estimation and Bayesian hierarchical inference are reviewed. It is shown that, with each novation in recursive methods, the limits of computability have expanded. The author argues that recursion is a vital aspect of human cognition, particularly in the development and interpretation of complex language. The paper also addresses the challenges of studying recursion and consciousness, such as the subjective nature of consciousness and the complexity of neural networks associated with conscious thought. Additionally, the paper examines the limitations of current theories of cognitive processing and language acquisition in understanding recursion and consciousness. The article concludes that investigating the relationship between recursion and consciousness is critical for developing a deeper understanding of language and cognitive processing. The author anticipates that a future recursion-based theory will help solve principal metaphysical conundrums of the past and the present.

Keywords: recursion; cognitive processing; consciousness; language acquisition; neural networks.

Philosophy is written in this all-encompassing book that is constantly open before our eyes, that is the universe; but it cannot be understood unless one first learns to understand the language and knows the characters in which it is written.

Galileo Galilei 'The Assayer'

The freedom of the will consists in the impossibility of knowing actions that still lie in the future.

We could know them only if causality were an *inner* necessity like that of logical inference.

Ludwig Wittgenstein '*Tractatus Logico-Philosophicus*'

The evolution of Turing computability up to post- and non-Turing models

The brief history of the concept of recursion and that of recursiveness of a function in their relation to the concepts of computation and computability was outlined in [1]. In what follows I will try to show that the present day of cognitive science is shaped by the meeting of general recursive (computational) approach therein with what may be labelled as the probabilistic turn in sciences of mind, most of all represented in various Bayesian approaches both in neurobiology and in Machine Learning. Taking into consideration Soare's recommendation to 'distinguish between the intensional meaning of Church's Thesis (that all effectively calculable functions are general recursive) versus Turing's Thesis (that all intuitively computable functions are computable by a Turing machine) [*italics omitted* — I. M.]' [2, p. 313], yet I will go on adhering to the extensional equality of the recursive and the computable, unless it undermines the overall idea of the paper.

As Soare reminds in [3, p. xxi–xxii], Turing introduced the concept of *oracle machines* in a one-page description in his dissertation, consisting of a Turing machine connected to an 'oracle' for querying during computation, which he compares to a modern local server connected to a large database. Emil Post expanded this concept and defined what he called *Turing reducibility*. A set B is Turing reducible to a set A ($B \leq_T A$) if there is an oracle machine that can compute B when the characteristic function of A is written on the oracle tape. This idea laid the foundation of modern computability theory. The notion helped define that of *Turing equivalence* as follows: two sets are Turing equivalent if each one is Turing reducible to the other. This is an equivalence relation, and it partitions the set of all sets of natural numbers into Turing equivalence classes. Turing equivalence can be used to compare the relative information content of non-computable (undecidable) sets. If sets A and B are Turing equivalent (i.e., $A \leq_T B$ and $B \leq_T A$), we view A and B as encoding the same information. Turing reducibility gives us a precise measure of the information they encode relative to other sets. The *Tur-*

ing degrees of these sets are equivalence classes containing sets with the same information content. Thus, the concept of Turing equivalence provides a precise and useful measure of the amount of information encoded in a non-computable set.

In [2, p. 302-303] Soare discusses the history of relative computability. The concept of relative computability is concerned with studying the relationship between two objects when one is computable relative to the other. Kleene and Post developed the concept of oracle Turing machines further by studying the computations that can be performed relative to an oracle. This led to the development of the theory of *degrees of unsolvability*, which is concerned with classifying computational problems based on their degree of difficulty relative to an oracle. In particular, Kleene and Post showed that there exist oracles such that the Turing machines relative to those oracles can compute functions that are uncomputable by any ordinary Turing machine.

In [4], under discussion is the problem of providing human-readable and understandable explanations of the decisions made by decision support systems using Bayesian networks, particularly in the context of the most probable explanation (MAP) problem. The concept of MAP-independence is introduced as a means of capturing the idea of relevance, or whether an unobserved variable might potentially impact the explanation. The computational complexity of several problems based on this concept is examined, particularly by addressing issues of specific kinds of Turing machines with an oracle. Oracle machines are instrumental in approaching computability of Bayesian inferences, specifically in the context of assessing the computational complexity of maximizing MAP-independence in Bayesian networks [4, p. 20].

Adding an oracle to a standard Turing machine as a solution to some otherwise intractable problems is also discussed in [5] in the context of modelling the 'autopoietic' human interaction with environment.

It seems significant for the topic of the probabilistic turn in cognitive science that the probabilistic Turing machine [6, p. 128-130] is often interpreted as a TM with a random oracle [7, p. 7]. The question of whether subsequent extensions of the classical (Turing) computability may help extend the class of computable functions is both historically and theoretically discussed in [8]. The same authors propose their own contribution to the required update of the classical construction in the form of Interactive Computation Theory [9]. It posits a simple model of interactive computing which uses a single component and an

environment that interact using input and output signals. The article describes a simple model of interactive computing which uses a single component and an environment that interact using input and output signals. The model builds on the theory of ω -automata and characterizes interactive computation in a stream setting.

The interactive computation model builds on the theory of ω -automata and characterizes interactive computation in a stream setting. ω -automata are machines that operate on infinite sequences of input. In the interactive model, the input is produced by the environment and, correspondingly, the output of the program is a sequence of actions in response to the input. However, the behaviour of the program is not determined purely by the input sequence, but also by the program state that changes over time as a result of its interaction with the environment.

The model of interactive computation is purely deterministic, as everything inside the system runs deterministically. The only non-deterministic aspects are any unpredictable time-delays that may occur due to external factors, such as the speed of the environment or input/output channels. But, as has been shown in Part I and will be confirmed further, empirically, probabilistic models of agent-environment interaction prove to be more realistic to the day.

Having said this, let us put our problem this way: if cognitive functions (say, language) are recursive, does it imply that the correlating neural functions (say, spike exchange) are recursive too? In Part I, this superstructure is casually labelled ‘meta-recursion’, which is not to be identified with the far more technical notion of *metarecursion* (a.k.a. *higher recursion*) as introduced by Gerald E. Sacks in [10; 11]. Probably, a better term could be ‘hyper-recursion’ or ‘super-recursion’, but let us admit that our intuitions on that matter has not yet matured enough to claim for a strict terminology.

Superstructures of recursion

As for language, there is an interesting study [12], that suggests that recursion might help the child acquire conceptual understanding of propositional attitudes and false beliefs. They argue that sentences are the appropriate medium for attitude concepts and claim that it is false complements rather than recursion per se that act as the ultimate trigger for the understanding of false beliefs. This is supported by ongoing language acquisition experiments. Children necessarily acquire sentential

complementation prior to acquiring the understanding of false beliefs, and truth contrasts enable children to recognize recursion more easily. This gives the child a way to represent false belief statements, which in turn leads to the understanding of false beliefs. Thus, to process propositions like ‘The girl said there was a butterfly in her hair (but it is a leaf)’ in order to be able to give the right answer to the question ‘What did the girl say there was in her hair?’ — ‘a butterfly’ instead of ‘a leaf’ — a child starts with detecting truth contrasts between clauses ‘the girl said’ (true) and ‘there was a butterfly’ (false) to proceed with both recursive complements in syntax and propositional attitudes in semantics.

In view of our aims, one can easily suggest that linguistic recursion is an ability acquired via learning, but it is made possible by recurrent loops in the brain. This relation may be characterized as that of weak supervenience, when a change in the ‘upper’ system necessarily correlates with *any* change in the ‘lower’ one, but *not the particular* change. This reservation is due to the brain plasticity and multi-functionality. Some intuitions into the brain’s recurrent structure and its computational performance may be obtained from studying recurrent artificial neural networks.

In [13] the computational and dynamical capabilities of biological neural networks are under discussion via comparison of the computational powers of diverse theoretical neural models with those of abstract computing devices. The authors suggest that some intrinsic computational capabilities of the brain might lie beyond the scope of Turing-equivalent models of computation, surpassing the potentialities of all current standard artificial models of computation.

They concluded that the computational powers of recurrent neural networks involved in the interactive computational framework follow similar patterns of characterization as those involved in a classical computational framework. The results showed that the incorporation of either evolving capabilities or some power of the continuum in a basic neural model provides an alternative and equivalent way towards the achievement of super-Turing computational capabilities. These achievements suggest that some intrinsic computational capabilities of the brain might lie beyond the scope of Turing-equivalent models of computation.

Recurrent neural networks are capable of processing dynamical information, which has been shown to be relevant to brain dynamics, such

as chaotic properties that cannot be described by the universal Turing machine model. Moreover, when equipped with real-valued weights, recurrent neural networks become capable of super-Turing computation, hence powerful enough to solve some hardest computable functions in polynomial time. These results were obtained by the authors through the so-called Thesis of Analog Computation, which states that no reasonable abstract analog device can be more powerful than first-order analog recurrent neural networks.

Reminiscent of the famous Marvin Minsky's metaphor of 'Society of Mind' is the fact that the power of interactive probabilistic computations reveals itself not only in neural networks, but in multi-agent systems as well. There is an interesting discussion of a framework for multi-agent reinforcement learning that adopts variational Bayes methods to approximate the opponents' conditional policies, to which each agent finds the best response and then improves their own policies [14]. Similar to the paper examined above, the reasoning studied here is called recursive reasoning because it represents the belief reasoning process where each agent considers the reasoning process of other agents, based on which it expects to make better decisions. Importantly, it allows an opponent to reason about the modelling agent rather than being a fixed type; the process can therefore be nested in a form as "I believe that you believe that I believe ...". Recursive reasoning is similar to the way that humans think about reasoning. Despite some initial trails, there has been little work that tries to adopt this idea into the multi-agent deep reinforcement learning setting. The methods are tested on both the matrix game and the differential game, and they are proved to converge in the self-play scenarios when there exists one Nash equilibrium.

Variational Bayes methods are used to approximate a complex probability distribution with a simpler distribution that is easier to calculate. In the context of multi-agent reinforcement learning, it is used to approximate the opponents' conditional policy. This approximation is then used by each agent to find the best response and improve their own policy. The paper discusses the framework of multi-agent reinforcement learning with action class and continuous state space. In this framework, each agent uses a deep neural network to approximate its policy and the value function. Moreover, to decentralize the learning process, they use the advantage actor-critic algorithm and the independent network learning algorithm. Finally, they use the variational Bayes method to approximate the opponents' conditional policy.

In this context, ‘conditional’ refers to the probability distribution of the action of one agent conditioned on the state of the environment and the policy of the other agent(s). The opponents’ conditional policy indicates how they would behave based on the current state of the environment and their own policies. The variational Bayes method is used to approximate this conditional policy.

In the experiments discussed in the paper, the proposed method is compared with other existing methods on two different benchmarks: one with continuous action space, and the other with discrete action space. The results show that the proposed method outperforms the existing ones on both benchmarks. Moreover, the authors also show the ablation study, where they remove the relevance of each component of the proposed framework in turn and compare the resultant methods. They argue that each component is critical to the overall performance of the algorithm.

In this study, “action space” refers to the set of all possible actions that an agent can take in a given environment. In a multi-agent reinforcement learning problem, the action space of each agent is dependent on the state of the environment and the policies of other agents. In the paper, the authors use two different benchmarks, one with continuous action space and the other with discrete action space.

General cognitive recursion

Michael Corballis is the author of numerous works developing the framework of the ‘recursive mind’. In his comprehensive book [15] he presents the mathematical concept of recursion by providing examples such as the factorial, Fibonacci sequence, and the Kanji script, and describes how recursion can extend indefinitely in theory but is limited in practice. Recursion is defined as a process that repeats itself and solves problems by breaking them down into simpler versions of the same problem. The author emphasizes that recursion gives rise to the concept of infinity, a mental achievement that is unique to humans.

He also discusses the features that distinguish human language from other forms of animal communication, specifically the concept of recursion. Recursion, the ability to embed structures within structures in a recursive fashion, allows humans to generate an infinite number of sentences. The author describes various forms of recursion, including center-embedded recursion, and provides examples of its use in language. He also debates whether animals have language and defines

language as a uniquely human trait. Finally, the author discusses the evolution of language, and how it evolved from manual gestures to vocalizations, before becoming the complex system of communication that it is today.

In Part 2 called ‘Mental Time Travel’, Corballis discusses the ability of humans to mentally transport themselves to other places and times, specifically through the power of memory and imagination. He introduces the concept of mental time travel, a phenomenon that allows humans to imagine and relive previous experiences, as well as plan possible future events. The author then connects mental time travel to language, suggesting that the ability to remember, plan, and tell stories may have been a critical factor in the evolution of human language and social cohesion. He provides several specific examples of mental time travel in everyday life, along with the role that storytelling and fiction play in this process.

In Part 3 — ‘Theory of Mind’, the author discusses how humans are able to understand what others are thinking or feeling, and introduces the concept of mind-reading, not as a psychic ability but as a natural ability to infer the mental perspectives of other people. He suggests that this ability is critical to social cohesion and cooperation and explains how theory of mind was also critical to the emergence of language. The chapter describes specific experiments that demonstrate the importance of theory of mind in social interactions, and how it develops in childhood. Additionally, the author discusses the limitations that may exist in our understanding of the thoughts and feelings of others, and the role that culture and socialization play in shaping our theory of mind abilities.

Corballis argues that recursion plays a critical role in the development of Theory of Mind. Recursive thinking allows individuals to infer not only what others are thinking or feeling but also to infer what others might think about what they are thinking or feeling. This capacity is essential for social interaction and communication, especially in the form of language. For example, in a conversation, people have to understand not only what the other person is saying but also what they mean by it, which requires recursive thinking. The author notes that theory of mind, like language, may have evolved from earlier cognitive capacities that were not language-based but that possessed recursive properties that allowed language to emerge.

Bayesian Brain

The growing family of Bayesian approaches in AI and other cognitive sciences uses various computational techniques to model cognition through anticipation. Of a principal interest for the present study is the method known as Recursive Bayesian Estimation. According to [16], RBE is an approach used in statistics and machine learning that estimates the current state of a system. It is used in robotics and automotive technology to perform a task that requires estimation. For example, a self-driving car can estimate its location using the Recursive Bayesian Estimation framework. The car obtains its starting position via a Global Positioning System (GPS), and its built-in algorithms then help it estimate its current location after a certain amount of time or distance. These algorithms often use the Bayesian mathematical concept in statistics.

The two essential concepts of Recursive Bayesian Estimation are recursive estimation and Bayesian inference. Recursive estimation refers to the process of estimating the current state based on the previous state, where the new state estimation will become the basis for the subsequent state estimation. Bayesian inference, on the other hand, makes it possible to obtain the state of an entity based on the prior state of an entity and the likelihood function or prediction.

Recursive Bayesian Estimation can be applied to other fields outside of statistics and machine learning. RBE applications range from the automotive industry to medical science. For example, the remaining useful life of medical devices can be predicted using recursive Bayesian estimation. Industrial robots are also designed to estimate their position using this estimation method. In general, recursive Bayesian estimation is a flexible mathematical approach that has helped advance technologies in many fields beyond statistics and machine learning.

Daunizeau in [17] explores a Bayesian approach to understanding the evolution of cognition in the brain. The author highlights the need to understand the principles of optimality in the brain to understand its functions. The concept also suggests that information theory offers a probabilistic account of how data is optimally processed, and Bayes' rule is crucial to understanding the brain.

The Bayesian Brain hypothesis identifies optimal solutions to the computational problems that the brain faces, promoting distinct properties such as precision, efficiency, flexibility, and sophistication. It takes a formal perspective on how selective pressure may shape the brain's

information processing mechanisms. It can be described on three levels of abstraction: the computational level, algorithmic level, and the physical/implementational level. It bridges the gap between the first two levels of analysis by describing how information processing should operate for realizing a given cognitive function.

Information theory provides a probabilistic account of how information is optimally transmitted and processed by relying on probability calculus and, more precisely, on Bayes' rule. It was originally developed to address problems of numerical signal processing such as data compressing and storage. It quickly extended beyond academic barriers, finding applications in statistics, artificial intelligence, physics, and telecommunications. More recently, information theory was successfully applied to model the way the brain perceives, understands and/or learns about its environment. Therefore, it helps explain how the brain perceives, learns, and understands information by providing a formal and consistent framework for introducing optimality principles into cognitive science research.

Optimality principles can help us understand the brain's mechanisms better. In cognitive science research, optimality principles are used to identify the best solution to a given computational problem that the brain faces while performing a cognitive function. Optimal solutions are defined in a formal and quantifiable manner, allowing researchers to make clear predictions about the expected performance of the cognitive system under different conditions. Furthermore, by introducing optimality principles into cognitive science research, it becomes possible to test whether the brain actually achieves the predicted performance or not, and in case it doesn't, whether it is due to the presence of a competing constraint or to some suboptimality in its mechanism.

The Bayesian Brain hypothesis is closely related to the concept of recursion. In general, recursion is a computational structure that allows the formulation of complex problems in terms of simpler ones. In other words, it allows for the chaining of a simple cognitive process to generate more complex behaviours. Similarly, the Bayesian Brain hypothesis proposes that cognitive functions at different levels of complexity can be understood in terms of Bayesian inference of probabilistic models. Given this, the Bayesian Brain framework can provide explanations of the cognitive processes that underlie recursive computation in perception, learning, and decision-making. In this sense, it constitutes a formal account of the type of recursive computation observed in the brain.

Recursion in brain computations is related to Bayesian hierarchical inference. Bayesian hierarchical inference refers to the process of inferring the structure of a complex generative model given a set of observed data. It involves assigning probabilities to the parameters and the hyperparameters of the model and using these probabilities to make inferences about the model's structure. Recursion refers to the computational structure of breaking down complex problems into simpler ones that can be understood as elementary building blocks, which can be further recursively nested to generate higher-order complexity. In the Bayesian Brain framework, the use of Bayesian hierarchical models enables the modeling of complex connectivity patterns at multiple levels of the nervous system. Using this approach, the brain is construed as generating more and more complex cognitive representations that exhibit the same type of recursive computational structure found in Bayesian hierarchical inference. Therefore, recursion in brain computations is related to Bayesian hierarchical inference, and this relationship provides a powerful means of explaining the brain's learning and computational mechanisms.

The Bayesian Brain Hypothesis is related to *Predictive Processing Theory*. PPT offers a bottom-up approach to perception, where top-down predictions about future sensory input modify the processing of incoming information. These predictions are based on generative models that implicitly encode statistical regularities within the external and internal environment. In contrast, the Bayesian Brain hypothesis provides a top-down approach to cognition, where cognitive processes are seen as implementing Bayes' rule to estimate the most likely cause of sensory input. Both models share the same Bayesian framework, which makes it possible to integrate sensory input and internal models to construct accurate perceptual representations of the world. In this sense, Bayesian models of the brain can provide a general theoretical framework for predictive processing, while PPT provides a mechanistic account of the precise neural computations that support Bayesian inference.

Recursive consciousness

According to the algorithmic information theory of consciousness [18], the existence and use of compressive models by cognitive systems in biological recurrent neural networks enables and provides the structure to phenomenal experience, which includes self-awareness as a part

of a better model. Recursion allows these cognitive systems to interact bidirectionally with the external world, giving rise to apparently complex (entropic but hierarchically organized) data. This theory is called KT, based on the mathematical theory of Kolmogorov complexity, which provides a natural framework to study and quantify consciousness from neurophysiological or neuroimaging data, given the premise that the primary role of the brain is information processing.

According to KT, the human brain creates models of its sensory inputs, which are essentially compressive models that generate outputs efficiently. These models are created in the form of biological recurrent neural networks with a hierarchical structure. The use of these compressive models allows the brain to perceive and interact with the external world, which provides the structure to the phenomenal experience, including self-awareness. Recursion refers here to the cyclic process of creating and refining these models to match the sensory inputs more precisely and perform more efficient compression.

A compressive model is a model of a system or a process that is capable of generating accurate outputs in response to particular inputs with high efficiency. In cognitive systems, these models are formed by the brain as a way of making sense of the sensory data it receives. The models are created using a hierarchy of neurons, and the connections between them are adjusted through a recursive process to better fit the inputs. The use of compressive models by cognitive systems in biological recurrent neural networks plays a vital role in the perception of the external world and provides the structural basis to the phenomenal experience, which includes self-awareness.

As mentioned above, the algorithmic information theory of consciousness presented in [18] is based on the concept of Kolmogorov complexity. Kolmogorov complexity provides a mathematical framework for quantifying the complexity of information. Kolmogorov complexity can conceptually split the Kolmogorov optimal program describing a data string into two parts: a set of bits describing its regularities and another that captures the rest (the part with no structure). The first term is the effective complexity, the minimal description of the regularities of the data. This notion provides the means to account for and separate regularities in data from noise. However, it is generally not possible to calculate the Kolmogorov complexity of an arbitrary string since we have no assurance that the Turing machine (TM) will halt. Despite this, within a limited computation scheme (*e.g.* in terms

of computation time or programming language resources), variants of algorithmic complexity, such as Lempel–Ziv–Welch compression, can be calculated. Lempel–Ziv–Welch compression is a simple yet fast algorithm that exploits the repetition of symbol sequences (one possible form of regularity), and its file length is equivalent to the entropy rate, which is an extension of the concept of entropy for stochastic sequences of symbols.

To connect issues of consciousness to those of language with respect to recursion, Roger Vergauwen, in his ‘Consciousness, Recursion and Language’ [19], examines the role of recursion in the human ability to generate and comprehend language. He argues that recursion is a key factor in the production and comprehension of complex linguistic structures and that it is closely related to the nature of consciousness. According to the author, recursion is an essential part of human cognition and allows for the creation of an infinite variety of novel linguistic expressions.

Vergauwen also discusses the relationship between recursion and consciousness. He argues that the ability to generate and process recursive linguistic structures is a key feature of conscious thought, and that it plays a critical role in the development of abstract thought. He further suggests that the recursive nature of language allowed humans to create symbolic representation and to engage in metacognition. Vergauwen also highlights some of the challenges associated with studying consciousness and recursion and discusses the limitations of current theories of cognitive processing and language acquisition.

Overall, Vergauwen highlights the importance of recursion in human cognition and the role it plays in generating and comprehending complex language. The author makes a case for the close relationship between recursion and consciousness, arguing that this connection is central to our understanding of the nature of human thought and the development of abstract concepts.

Vergauwen discusses several challenges associated with studying consciousness. One of the main issues is the subjective nature of consciousness. Consciousness is a first-person experience that cannot be directly observed or measured by external means. This makes it difficult to develop objective criteria for studying consciousness and to determine what aspects of consciousness are relevant for linguistic processing.

Another significant challenge is the complexity of the neural net-

works involved in conscious thought. Consciousness arises from the complex interaction of neural processes across multiple regions of the brain. This makes it difficult to isolate specific brain areas or mechanisms that are responsible for generating recursive structures in language. Similarly, it is challenging to identify the specific neural mechanisms or structures that are specifically associated with consciousness.

Finally, Vergauwen notes that the recursive nature of language creates a kind of ‘infinite regress,’ which poses theoretical challenges for understanding how the brain can process and generate infinitely recursive linguistic structures. This makes it challenging to investigate the neural underpinnings of recursion in language and to develop clear models of how it is processed in the brain. Despite these challenges, the author argues that the study of consciousness and recursion is a critical endeavour for understanding the nature of human thought and for developing a deeper understanding of language and cognitive processing.

According to Vergauwen, current theories of cognitive processing and language acquisition have several limitations when it comes to understanding recursion and consciousness. One limitation is that existing models of linguistic processing often assume a modular and hierarchical organization of the brain, which may not be fully supported by empirical evidence. Hierarchical models of the brain may struggle in accounting for the recursive nature of language processing, especially in cases when recursion is nested infinitely.

Another limitation is that existing theories of language acquisition are often based on the idea of a specific language module, which may not fully capture the complexity of language processing. This specific language module hypothesis may fail to account for the fact that language processing is deeply intertwined with other cognitive processes, such as perception, attention and memory. This may lead to models which do not fully consider the joint effects that human cognitive processes have on language processing.

Finally, Vergauwen points out that current theories of linguistic processing are limited by their over reliance on behavioural and neuro-imaging evidence. These restrictions make it difficult to determine the specific neural mechanisms that support the generation and comprehension of recursive structures in language. Vergauwen highlights the need for a multidisciplinary approach that integrates behavioural evidence and neuroscientific evidence from other species, including non-human primates or rodents, to guide our understanding of language processing.

Giulio Tononi with co-authors has presented *Integrated Information Theory (IIT) of consciousness 3.0* [20] that explains how physical mechanisms, such as neurons or logic gates, must be configured to generate experience (phenomenology). The theory defines intrinsic information as ‘differences that make a difference’ within a system, and integrated information as information specified by a whole that cannot be reduced to that specified by its parts.

According to Integrated Information Theory (IIT), consciousness arises as a result of the causal relationships between a system of mechanisms that are highly integrated and differentiated. The theory posits that conscious experiences are generated by a system with maximum integrated information, or highest Φ value, which characterizes the degree to which the system’s parts are mutually dependent and share cause-effect information. A system with high Φ is capable of generating an experience.

Intrinsic information is the information generated by a mechanism over and above the information generated by its individual components, while integrated information is the conceptual information generated by a system of elements over and above the information generated by its minimal parts.

According to Integrated Information Theory (IIT), an experience is identified by a complex of concepts that describe the quality of the experience itself, how much of it there is, and what it is like. These concepts are integrated together among the various parts of the complex, and the more integrated they are, the more unified the experience is. The identity of an experience is then taken to be one complex of concepts, which is irreducible to or different from the other complexes that can be defined by other conceptual decompositions.

Recursion is a central concept in Integrated Information Theory (IIT), which holds that consciousness arises as a result of a system of mechanisms and their causal relationships: ‘The idea that “feed-back”, “reentry”, or “recursion” of some kind may be an essential ingredient of consciousness has many proponents’ [20, p. 19]. As described in the article, a complex is considered a mechanism if it generates cause-effect repertoires that are irreducible to those produced by independent parts of the system. These mechanisms can then be combined into larger, more integrated mechanisms, which can be iteratively decomposed into smaller mechanisms again. This recursive organization of mechanisms and their causal relationships is critical for understanding how con-

consciousness arises from the interactions between different parts of the nervous system.

Another recursive account of consciousness [21] defines it as a recursive, spatiotemporal self-location. Consciousness arises as a singular, unified field of recursive self-awareness with explicit orientational characteristics. The psychological structures supporting self-located subjectivity involve an evolutionary elaboration of the two basic elements necessary for extending self-regulation into behavioural interaction with the environment. Understanding consciousness starts with the recognition that cognitive systems serve the biological self-regulatory regime in which they subsist.

The primary reference frame of active waking cognition is *current state* (CS), which is a computation of spatiotemporal self-location updated moment to moment from self-movement feedback (as well as anticipatory *next perceptual state* (FM) of expected feedback). The two basic elements that support self-located subjectivity are an orientative reference frame that structures ongoing interaction in terms of controllable spatiotemporal parameters, and processing architecture that relates behaviour to homeostatic needs via feedback. These two elements are essential for extending self-regulation into behavioural interaction with the environment.

According to the theory, the dedication of cognitive architecture to an anticipative feedforward processing format for speed and energy efficiency implies that the recursive circuitry required to underwrite conscious subjectivity is most likely to have evolved out of such predictive cognitive architecture. Specifically, the theory proposes that evolutionary pressures for ever more energy-efficient sensorimotor processing have encouraged the progressive attenuation of feedforward processing circuitry into a simplified recursive feedforward circuit capable of underwriting autoreferential conscious self-awareness.

The article proposes that consciousness arises through a recursive, spatiotemporal, self-location event schema functioning as a 'self-activated, recursive, working memory circuit' [21, p. 417]. Recursion is an important part of this proposal because the recursive circuitry is what underlies conscious subjectivity by processing anticipatory feedforward predictions together with feedback-derived self-location information to generate and maintain a spatiotemporal model of the self and the environment. Recursion is also discussed as a related concept to hierarchy and attenuation, as it is thought that recursive

circuitry emerges out of the attenuation of more complex hierarchical feedforward processing.

Conclusion

The overview of the topic I have managed to present above shows, as it seems, that the initially purely mathematical notion, attempted to be formalized by the greatest minds of the discipline and finally stranded on the shore of ambiguities somewhere between induction and computation, nevertheless represent some important intuition concerning automated enumerating and counting, computations in their broadest sense, language acquisition and mastering and, finally, mind and consciousness. This implies that, being jack of all trades, recursion has no specific traits to define any of these in particular — maybe except for computation, which is disputable, too. But if, as I have conjectured, the recursiveness of all these capabilities is ensured by recurrent loops in the brain, upon which it is weakly supervenient, we have a chance to approach a possible general theory that will explain away all the mentioned capabilities as ‘folk-psychological’ hypostases of one important capability of computational optimization. Taking into consideration, besides other things, that recurrence is usually taken for a special case of recursion.

This future comprehensive theory will help us crack the code, with which philosophy is inscribed in the book of Universe, and which will lift the veil on our actions lying ahead rejecting ‘the absurd tale about free will’ (V.I. Lenin).

References

1. Mikhailov I.F. The Concept of Recursion in Cognitive Studies. Part I: From Mathematics to Cognition // Philosophical Problems of IT & Cyberspace (PhilIT&C). 2024. № 1. P. 58-76.
2. Soare R. I. Computability and Recursion // Bulletin of Symbolic Logic. 1996. Vol. 2, № 3. P. 284-321.
3. Soare R. I. Turing Computability: Theory and Applications. Berlin Heidelberg: Springer-Verlag, 2016. 289 p.
4. Kwisthout J. Motivating explanations in Bayesian networks using MAP-independence // International Journal of Approximate Reasoning. Elsevier Inc., 2023. Vol. 153. P. 18-28.
5. Kato T. et al. Explore/exploit strategic movement of co-occurrence network based on a conscious/subconscious model // 2015 54th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE). IEEE, 2015. P. 301-306.
6. Arora S., Barak B. Randomized computation // Computational Complexity. 2012. P. 123-142.

7. Antonelli M., Lago U.D., Pistone P. On Randomized Computational Models and Complexity Classes: a Historical Overview, 2024.
8. Van Leeuwen J., Wiedermann J. The Turing Machine Paradigm in Contemporary Computing BT — Mathematics Unlimited — 2001 and Beyond / ed. Engquist B., Schmid W. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001. P. 1139-1155.
9. Van Leeuwen J., Wiedermann J. A Theory of Interactive Computation BT — Interactive Computation: The New Paradigm / ed. Goldin D., Smolka S.A., Wegner P. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. P. 119-142.
10. Sacks G. E. Metarecursion Theory // Studies in Logic and the Foundations of Mathematics. Elsevier, 1967. Vol. 46, № C. P. 243-263.
11. Sacks G. E. Higher recursion theory. 2nd ed. Cambridge, UK: Cambridge University Press, 1990.
12. De Villiers J., Hobbs K., Hollebrandse B. Recursive Complements and Propositional Attitudes // Recursion: Complexity in Cognition. Studies in Theoretical Psycholinguistics, vol. 43 / ed. Roeper T., Speas M. Springer, Cham, 2014. P. 221-242.
13. Cabessa J., Villa A. E. P. Recurrent Neural Networks and Super-Turing Interactive Computation // Artificial Neural Networks. Springer Series in Bio-/Neuroinformatics, vol. 4. Springer, Cham., 2015. P. 1-29.
14. Wen Y. et al. Probabilistic recursive reasoning for multi-agent reinforcement learning // 7th International Conference on Learning Representations, ICLR 2019, 2019. P. 1-20.
15. Corballis M. C. The Recursive Mind. The Origins of Human Language, Thought, and Civilization // The Recursive Mind. Princeton University Press, 2014. 309 p.
16. TECHSLANG. What is Recursive Bayesian Estimation? A short definition of Recursive Bayesian Estimation [Electronic resource]. 2024. P. 1-5. URL: <https://www.techslang.com/definition/what-is-recursive-bayesian-estimation> (accessed: 22.02.2024).
17. Daunizeau J. The Bayesian Brain: An Evolutionary Approach to Cognition // Encyclopedia of Behavioral Neuroscience: Volumes 1-3, Second edition. 2021. Vol. 1-3. P. V2-202-V2-221.
18. Ruffini G. An algorithmic information theory of consciousness // Neurosci Conscious. 2017. Vol. 2017, № 1.
19. Vergauwen R. Consciousness, Recursion and Language // Language and Recursion. NY: Springer New York, 2014. Vol. 9781461494. P. 169-179.
20. Oizumi M., Albantakis L., Tononi G. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0 // PLoS Comput Biol / ed. Sporns O. Public Library of Science, 2014. Vol. 10, № 5. P. e1003588.
21. Peters F. Consciousness as recursive, spatiotemporal self-location // Psychol Res. 2010. Vol. 74, № 4. P. 407-421.

УДК 159.95, 519.7

DOI 10.17726/philIT.2024.2.2



Аналогия как базовая функция мышления

Александр Хомяков,
магистр физических наук
Санкт-Петербург, Россия
alexander.xom@gmail.com

Аннотация. Автор проводит аналитическое исследование, в котором выдвигается гипотеза о роли аналогии в мышлении. В статье представлена критика гипотезы структурного отображения в аналогии как объяснения аналогичного аргумента и будет предложено альтернативное объяснение — функциональная гипотеза аналогичного аргумента. Также будет показано, как при помощи разработанного метода аналогии могут реализовываться такие функции мышления, как распознавание и память, метафора и силлогизм, обобщение и онтология, дедукция и индукция. Отстаивается гипотеза, что эти разные функции есть не более чем разные проявления одного механизма распространения схем мышления через аналогию. В предыдущей нашей работе [1] мы описали результаты экспериментов по получению аналогов слов на основе статистики предикатов, извлеченных из текстов простым парсером. В данной статье, продолжая тему аналогии, мы проанализируем два аспекта: как находятся аналогичные выражения (аналогичный аргумент) при решении задач и обучении; какова роль аналогии в мышлении, а именно: какие процессы мышления реализуются при помощи аналогии. В статье критически рассмотрены наиболее известные подходы к объяснению и попыткам реализации аналогии. Нами предлагается другая гипотеза, называемая функциональной аналогией, которая отличается простотой объяснения аналогичного аргумента, большим соответствием примерам аналогии и программной реализации.

Ключевые слова: теория аналогии; когнитивная лингвистика; дистрибутивная семантика; аналогичный аргумент; мышление; метафора.

Analogy as a basic function of thinking

Alexander B. Khomyakov,
Master of Physical Sciences
Saint-Petersburg, Russia
alexander.xom@gmail.com

Abstract. We present an analytical study that puts forward a hypothesis about the role of analogy in thinking. The article presents a critique of the structural mapping hypothesis in analogy as an explanation of the analogical argument and will offer an alternative explanation — the functional hypothesis of the analogical argument. It will also be shown how such functions of thinking as recognition and memory, metaphor and syllogism, generalization and ontology, deduction and induction can be realized with the help of the analogy method developed by the author. The article puts forward a hypothesis that these different functions are nothing more than different manifestations of one mechanism for the dissemination of thinking patterns through analogy.

In the previous article [1] we described the results of experiments on obtaining analogs of words based on the statistics of predicates extracted from texts by a simple parser. In this article, continuing the topic of analogy, we will analyze two aspects:

- how similar expressions (analogous argument) are found when solving problems and learning;
- what is the role of analogy in thinking, namely, what thinking processes are realized with the help of analogy.

The article critically examines the most well-known approaches to explaining and attempting to implement analogy. The authors propose another hypothesis, called functional analogy, which is distinguished by the simplicity of explaining the analogical argument, greater correspondence to examples of analogy, in the authors' opinion, and the possibility of programmatic implementation.

Keywords: analogy theory; cognitive linguistics; distributive semantics; analogical argument; thinking; metaphor.

Введение

Роль аналогии в мышлении является важной областью изучения в современной психологии, особенно в понимании когнитивных процессов, решении проблем и обучении. Аналогическое рассуждение, как способность определять соответствия между различными сущностями на основе общих отношений, является основополагающим для человеческого познания. Оно облегчает получение знаний, решение проблем и объяснение новой информации в различных контекстах, включая образование, науку и повседневную жизнь.

Для понимания аналогии было предложено несколько теоретических основ. Теория структурно-картографического отображения Д. Гентнер [5], которая на сегодня является наиболее популярной,

утверждает, что аналогическое рассуждение включает в себя отображение структуры отношений одной области на другую, целевую, подчеркивая реляционные сходства над сходством атрибутов. Известна также теория множественных ограничений, разработанная К. Холиок и П. Тагардом [13]. Этот подход объединяет различные ограничения (структурные, семантические и прагматические), которые влияют на то, как аналогии формируются и используются в рассуждениях. Или, например, модель AMBR Б., разработанная Б. Кокиновым и А. Петровым [4]. В ней выстраивается аналоговое рассуждение на основе памяти.

Но ни она из теорий не может быть подтверждена или опровергнута эмпирически, так как для аналогии трудно найти нейронные корреляты в мозгу. И поэтому некоторые исследователи сосредоточились на реализации теорий в виде программного алгоритма, которые должны повторять аналогичные рассуждения человека. Такие попытки предпринимались авторами двух теорий. Это программа SME, созданная Д. Гентнер в 1985 году и продолженная ее последователем К. Форбусом [2], и программа СоруСат [8], созданная Д. Хофштадтером в 1988 году, работу над которой ведет в настоящее время М. Митчелл. Рассмотрим подробнее эти две теории аналогии.

Материалы исследования

Наиболее популярным является подход Дедре Гентнер, выдающегося когнитивного психолога, в настоящее время профессора психологии в Северо-Западном университете США. Она известна своей работой по аналоговому рассуждению, в частности благодаря разработке теории структурного отображения. Эта теория утверждает, что аналогии формируются путем отображения реляционных структур между базовой областью (знакомой ситуацией) и целевой областью (новой ситуацией). Исследования Д. Гентнер подчеркивают, как эти отображения облегчают обучение и рассуждение, позволяя людям делать выводы и генерировать новые знания из предыдущего опыта.

Кеннет Форбус в своих исследованиях в соавторстве с Д. Гентнер развивает вычислительные модели, которые имитируют человеческое аналогичное мышление. Эти модели направлены на воспроизведение когнитивных процессов, задействованных в проведении аналогий. Он демонстрирует, что предлагаемая им про-

грамма аналогии обучается знаниям не на тысячах примеров, как в нейросетевых подходах, а на десятках. К. Форбус не опубликовал никаких конкретных математических расчетов для поиска аналогий. Однако он разработал вычислительные модели и алгоритмы для аналогических рассуждений, которые не опираются на явные математические формулы. Он использует структурированные представления, такие как исчисление предикатов и концептуальные графы, для фиксации реляционной структуры концепций и ситуаций.

Взгляд Дугласа Хофштадтера [12] на аналогию существенно отличается от других теорий в нескольких ключевых аспектах, особенно в том, как он концептуализирует роль и природу аналогии в познании. По его мнению, аналогия — это не просто инструмент для рассуждения или решения проблем, но и фундаментальная основа всех когнитивных процессов. Д. Хофштадтер утверждает, что «практически каждая мысль» — это аналогия, предполагая, что сами концепции — это пучки аналогий. Это контрастирует с теориями, которые рассматривают аналогию как специализированную форму рассуждения. Д. Хофштадтер критикует тенденцию определять аналогию в строгих, формальных терминах, таких как «А относится к В, как С относится к D». Он считает, что такие определения чрезмерно упрощают сложную, текучую природу аналогового мышления. Д. Хофштадтер вводит идею восприятия «глубинной сущности» концепций посредством аналогии, которая подразумевает распознавание базовых сходств, которые могут быть не очевидны сразу. Это расходится с другими теориями, которые могут отдавать приоритет наблюдаемым признакам или структурным соответствиям в аналоговом рассуждении. Д. Хофштадтер полагает, что такие модели не в состоянии охватить богатство и изменчивость человеческих мыслительных процессов, которые по своей сути нелинейны и динамичны.

Идеи Д. Хофштадтера продолжает его ученица Мелани Митчелл, профессор в Институте Санта-Фе. Вместе они создали программу CopyCat [8], которая, в отличие от SME, основывается на статистическом и вероятностном подходе. CopyCat решает задачи аналогий в ограниченной области, например преобразование строк букв (например, «abc» к «abd»). Программа должна определить аналогичное преобразование для другой строки (например, «ijk» к «?»), что требует «понимания» структуры и правил преобразова-

ния. Но пока программа не может быть распространена на более сложные задачи из-за высокой вычислительной сложности. Основным элементом программы является Slipnet, который функционирует как долговременная память программы, где расстояния между концептами могут изменяться в процессе работы программы. Чем меньше расстояние между концептами, тем более вероятно, что они будут сопоставлены друг с другом в процессе аналогии.

Анализ рассмотренных гипотез аналогии

Когнитивные механизмы, лежащие в основе аналогического мышления, согласно гипотезе Д. Гентнер, реализуются в ходе следующих этапов:

1. Извлечение: способность извлекать из памяти соответствующие знания, которые могут служить источником для проведения аналогий.

2. Картирование: включает в себя определение и согласование отношений между исходными и целевыми доменами, что имеет решающее значение для эффективного использования аналогии.

3. Передача: процесс применения знаний, полученных с помощью аналогии, к новому контексту, облегчающий обучение и решение проблем.

Проблема у такой гипотезы начинается с первого же этапа. Каким образом извлекается именно то знание, которое уже подходит для создания аналогии? Если исключать простой перебор, это может быть только если аналогия уже «предвидится» на этапе поиска знания, а значит, она уже должна быть, что противоречит всему подходу поэтапного формирования аналогии. На втором этапе встает вопрос, как происходит согласование исходного знания с целевым по отношениям, если они выражены разными, но аналогичными лексемами отношений (предикатами).

Существенная проблема обсуждаемого подхода в том, что непонятно, как выбираются отношения для сходства. К. Форбус не указывает конкретный метод, что свидетельствует о том, что это перебор, а значит, почти нерешаемая задача, если надо перебрать комбинацию, например, по три отношения для сравнения из всего набора возможных отношений в языке.

Еще одна проблема реализации указанных подходов и прежде всего подхода Гентнер-Форбуса состоит в том, что она предполагает ручную разметку отношений в тексте. Именно поэтому такой

подход, который был впервые описан в 1985 году в работе Гентнер, до сих пор не получил реализации в компьютерных системах, в отличие от нейросетевых подходов, зародившихся примерно в то же время и получивших недавно возможность «обучения без учителя» на текстах с маской.

Также непонятно, сколько должно быть найдено отношений, чтобы структурная аналогия «состоялась» в подходе Гентнер-Форбуса. Одно или несколько, и каких именно, ничего не говорит. В приводимых авторами гипотезы структурного подхода к аналогии примерах их несколько — от 2 до 5. В предлагаемом нами методе поиска аналогии это может быть от 10 до 100 предикатов отношений, которые определяют аналогию, отобранную автоматически без ручной разметки графов. Это принципиальное отличие в возможности реализации и надежности полученной аналогии.

Проблемой в подходе Гентнер-Форбуса также является то, что в нем не могут учитываться синонимы лексем, чтобы использовать по-разному выраженные, но одинаковые по смыслу в данном контексте отношения, что не позволяет более широко использовать сходство по отношениям. Получается так, что метод структурного отображения аналогичного аргумента страдает от отсутствия аналогии между лексемами, которые в теории Гентнер-Форбуса не представлены.

Трудность в теории Д. Гентнер можно пояснить на простом примере. Одним из способов представления системных отношений являются таблицы, как в таблице Менделеева, которые устанавливают отношения между сущностями вида «ниже — больше» или «левее — меньше». Но таким образом могут быть отражены совершенно разные явления и это не значит, что между ними есть аналогия, кроме формы таблицы. Или надо признать все таблицы аналогами. Представляется, что нужны еще какие-то дополнительные связи между доменами, как у времени с часами (часы показывают время) или у пути с жизнью (всю жизнь куда-то идти), которые придают целевому явлению дополнительные свойства. Например, начала и конца жизни (перенос от начала и конца пути на местности).

Функциональная аналогия

Предлагается другое объяснение для формирования аналогичного аргумента. Мы находим общий принцип в новом и известном

явлении, чтобы применить его как модель нового с предсказанием свойств в новом явлении. Только движение по кругу без видимой связи обуславливает аналогию солнечной системы и атома. То же самое с электрическим током по аналогии с течением воды в трубах, энергия (работа) которых зависит от приложенного давления или напряжения (есть еще сходство по форме трубы и провода). Но на этом аналогия заканчивается, так как по проводам ничего на самом деле не течет и даже энергия передается не по самим проводам, а по направляемым ими полям вокруг, а электрон не имеет формы шара, как планета. Мы назвали это функциональной аналогией. Это сходство по действию, функции, что видно по народной аналогии — хитер как лиса, опасен как волк. «Как» обозначает действие «Как это сделано?». И для такой аналогии достаточно одного предиката в обобщенном через аналогию виде, а не множества как в структурном отображении. Наиболее близко к этому применение одинаковых математических уравнений для разных процессов. Сходство отношений, взаимодействия элементов аналогии есть сходство функциональное. И это одно-единственное основание сходства.

Наблюдая новое явление, как, например, структуру атома, мы определяем обобщенное, абстрактное отношение «вращается вокруг центра *без видимой связи*». И именно этим свойством «без видимой связи», которое выявлено в опытах с атомами, определяется выбор аналогии планетарной системы как наиболее распространенной аналогичной схемы из всех. В чем отличие?

Во-первых, из цели определяется обобщенное отношение «вращается вокруг». Именно обобщенное, для того чтобы расширить поиски аналогии. Это функциональное отношение. Такие отношения присутствуют во многих процессах, например карусель, воронка воды, центрифуга, планетарная система.

Во-вторых, из всех возможных «вращений» избрана именно планетарная система благодаря условию «без видимой связи», которая обнаруживается в опыте над атомами. Именно это свойство отбрасывает другие варианты. Это схема «найти обобщающую схему, затем найти специализирующий признак, чтобы создать новую схему». По этой же схеме происходит обучение и создание новых схем, о чем еще будет подробнее изложено ниже. Мы называем такую последовательность функциональной аналогией. И, в отличие от гипотезы Гентнер, для поиска аналогии в ней не

требуется совокупности несколько отношений как структурного картирования, но требуется учет свойств и атрибутов, определяемых в целевом явлении.

Аналогия в мышлении

Напомним, что все дальнейшие рассуждения основаны на предложенном нами методе получения аналогов методом дистрибутивной семантики, применяемой к предикатам, выделенным автоматически из текста методом *predicatIDF*. Это позволяет получать аналоги разной степени общности в зависимости от числа общих предикатов, которые берутся для рассмотрения. При максимальном числе предикатов (до 100) выделяются синонимы и омонимы. При минимальном — одинаковые по какому-то одному предикату отношения или признака лексемы. Это общности вида «нечто зеленое», «живое», «развивающееся» с субстантивами «зелень», «животное», «развитие». Надо отметить, что предикаты лексем отражают не только их признаки, но и латентные отношения с другими лексемами, что является более надежным признаком аналогии. Коротко говоря, если объект, обозначаемый лексемой, ведет себя в разных ситуациях так же, как и другой объект, то обозначающая его лексема является аналогом, но в разной степени, что зависит от числа общих предикатов.

Хотя в предыдущей статье аналогия рассматривалась только как лексическая функция, предложенный метод не ограничивается только аналогиями лексем. Предикативная структура как таковая присуща нашему мышлению априори, все, чем мы оперируем в мышлении, имеет предикативный строй. Например, в визуальной модальности предикация связана с перемещением глаз в разных направлениях и условно ее можно изобразить так: если наложить на любое изображение сетку как в тетрадке в клетку, то любой квадратик картинки будет относиться к другому как «красный на один квадратик ниже, выше, направо или налево от синего». Это предикаты движения глаз, и таким образом, через такие предикаты перемещения, можно описать любую картинку, в зависимости от разрешения сетки. И такое предикативное разложение имеет некоторую ассоциацию (не осмелимся сказать большего) с выявленными в гиппокампе нейронами решетки и нейронами места. Именно нейроны решетки служат для создания предикативных отношений между нейронами мест при перемещении от одного к другому (ра-

нее была выявлена связь нейронов решетки с перемещением подопытного животного). Так что такая предикативная структура изображений может создавать аналогию мест на картинке, что в итоге может складываться в ощущение одного (близкого) цвета, текстуры и границы (аналогично расположенные по одному предикату и разные по-другому).

Теперь проанализируем возможную роль аналогии в известных нам операциях мышления, а именно распознавании, сравнении, силлогизме, метафоре, онтологии, дедукции и индукции, обучении. В работах упомянутых ранее авторов рассматривается только общая роль аналогии в распознавании и обучении, в создании новых гипотез, в метафоре, без определения того, как они осуществляются. Ниже будет показано, как эти и другие известные способы мышления реализуются при помощи аналогии.

Распознавание

Основной проблемой распознавания является усматривание схожих признаков объектов, которые в деталях могут сильно различаться. Нейросети распознают за счет того, что в процессе обучения им показывается очень много объектов класса и при небольшом обобщении между набранными разнообразными признаками (сглаживании) они способны распознавать объекты, немного различающиеся с представленными при обучении, но того же класса. Основным недостатком такого обучения в том, что для этого нейросети требуется много примеров для постепенного смещения весов (параметров) нейросети в нужном направлении. Человек учится явно быстрее, что мы обсудим далее в разделе об аналогии в обучении.

Аналогия объединяет несколько схожих лексем и образов как близких по предикатам с учетом избирательности и частотности, что отмечено в нашей предыдущей статье [1]. И распознавание происходит на основе распознавания предикатов нового объекта. Если он имеет такие же предикаты, то это лексема и объект, уже известные нам. Именно так мы относим их к одному классу, то есть распознаем. Важно отметить, что распознавание осуществляется не за счет обучения на большом числе примеров. Это комбинация предикатов, которые, как примитивы, уже известны и просто собираются в новую совокупность, аналогичную известной по классу. И мы называем новое так же, как известное (имя класса). Это свой-

ство композиционности будет рассмотрено далее.

Здесь следует привести пример, который наглядно поясняет, как происходит распознавание по аналогии. Все цвета являются аналогами друг другу как цвета. Они хорошо выделяются как аналоги в предложенном методе *predicatIDF*. Так вот, все вместе они образуют понятие цвета (цвет происходит от старославянского *květъ*, обозначающего не только цветок, но и луг, где много разноцветных цветов, поэтому «цвет» выбран как обобщающее понятие окраски в славянском языке). Так же любой выбранный цвет сразу ассоциируется с линейкой других цветов, которые входят в аналогию. Получается, что присвоение какого-то одного цвета предмету позволяет быть готовым распознать так же такие же предметы всех других цветов, так как в его композицию предикатов входит один цвет, а у него есть в качестве аналогов все другие. Это важное свойство композиционности аналогов, создающее возможность распознавания даже таких предметов, которые обладатель мышления никогда не видел, что позволяет учиться не на миллионах примеров, а даже на одном. Такие линейки возникают везде как перечисления аналогичного — чисел, имен людей, видов транспорта и др., создавая объем понятий, которые имеют указанное обобщающее название (имя понятия).

При этом аналоги формируются динамически, с распознаванием новой композиции предикатов как предмета возникают новые предикаты и аналоги с их участием, что создает новые возможности для распознавания и это реализует основной механизм познания, частью которого распознавание, собственно, и является.

Схема или композиция

В связи с упомянутой композиционностью как способом определения предмета или понятия (а любое слово — это обозначение не единичного предмета, а общности всех возможных аналогичных предметов), стоит определить понятия композиции, или схемы, предмета. Сочетание предикатов как описание предмета через аналогию лексем определяет все аналогичные предметы, как показано выше, при распознавании. Но это не случайное и единичное сочетание. Это именно обобщение сочетаний множества близких по аналогии лексем в одинаковой синтаксической конструкции. Схема предмета таким образом является наиболее частотным сочетанием, композицией аналогичных лексем в определенной син-

таксической конструкции как решетке (сетке). Хотя возможны различные способы описания предмета, сходство между которыми менее очевидно и требует отдельного анализа, сейчас мы говорим о схеме как совокупности аналогичных лексем, наиболее часто встречающихся вместе в совокупности случаев. Таким образом, мы будем называть схемой или композицией А все случаи сочетаний аналогичных лексем с одинаковыми синтаксическими связями.

Сравнение

Сравнение нечасто встречается в научной литературе как процесс мышления. В то же время весь процесс мышления пронизан оценкой ожидания и результата каждого действия, каждой мысли. Также сравнение является таким же приемом для логического вывода, как и силлогизм, только в форме «если (сравнение), то (вывод)». Например, если Вася придет раньше Пети на финиш, то он займет первое место. Вася пришел в 10.30, а Петя в 10.35.

Сравнение всегда реализуется через оператора отношения сравниваемых величин. Отношения в линейках могут быть упорядоченными, как, например, больше, дальше, вместительнее, так и неупорядоченными, то есть отношение при сравнении обозначается как «разное, другое», например имя. Сами сравниваемые лексемы (и понятия, выражаемые ими, далее мы не будем различать эти слова), чтобы быть сравнимыми, должны быть аналогами друг другу и входить в описанные выше линейки аналогичных понятий. Иначе они несравнимы. То есть аналогия является условием сравнения.

Приведем простой пример для пояснения. Есть две композиции. Вася имеет рост 178 см, а Петя имеет рост 176 см. Имена Вася и Петя являются аналогами в понятии «имя человека», величины 176 см и 178 см тоже являются аналогами в понятии «измерительная линейка». Связь человека и величин обозначена понятием «роста», которое является предикатом к понятию «человек» (обобщение имен). И только наличие общего понятия позволяет сравнивать по росту. В сравнении определяется разница в росте аналогичных имен. И делается вывод отношением имен «Вася выше Пети, так как 178 больше 176 в измерительной линейке». Но чтобы сравнить имена, нам надо перейти от понятия роста к другому понятию — мальчик (человек), обладающему разными именами. И тогда мы

сможем в сравнении сказать «мальчик Петя и мальчик Вася имеют разные имена». Такова роль общего понятия для сравнения: схема сходства (рост, мальчик) позволяет сопоставлять композиции по различающимся частям (величины или имени).

Силлогизм и Метафора

Силлогизм — это вывод из двух посылок (пропозиций) третьей. Мы же рассматриваем силлогизм как последовательность «посылка — аналогия — вывод», то есть как рассуждение по аналогии, как она была описана выше. Если у вас уже есть известная пропозиция и есть аналогия с неизвестным, к которому надо установить пропозицию, это реализуется через указанную последовательность из трех утверждений, среднее из которых не что иное, как установление аналогии между средним и меньшим терминами силлогизма. Поясним для ясности на классическом примере силлогизма про Сократа.

«Человек смертен» — это исходная пропозиция, которая нам известна. «Сократ — человек» — это и есть установление факта аналогии между именем Сократ и понятием Человек, в объем которого входит имя Сократ. Замена «человек» в исходной пропозиции на «Сократ» в силу аналогии приводит к новой искомой пропозиции «Сократ смертен». Аналогичный силлогизм мог бы выглядеть так, что более естественно для обыденных рассуждений. «Петя смертен, а Вася что Петя, тоже смертен». И тут утверждение «Вася что Петя» утверждает аналогию между ними как людьми.

Эти примеры показывают сходство силлогизма с метафорой по своей структуре. В теории метафоры принято говорить об утверждении-источнике и утверждении-цели, преобразование источника в который происходит за счет установления связи членов утверждений через предикат.

Возьмем простой пример. «Часы идут» — это утверждение-источник, или известное. «Часы показывают время» — это предикат, который устанавливает связь между известным понятием источника (часы) и новым понятием времени, свойство которого по аналогии надо определить. В итоге его установления получается метафора «Время идет» по той же схеме, что и в силлогизме. Отличие заключается в том, что здесь аналогия устанавливается по одному предикату, указывающему скорее на ассоциацию часов со временем, что можно считать слабой аналогией. Но она свой-

ственна всем новым понятиям, для которых еще не распространены многочисленные предикаты, по которым можно было бы установить сильную аналогию, как в случае силлогизма. Но по сути это тот же процесс установления вывода из известного утверждения по аналогии, которая установлена по указанному предикату (в силлогизме он не указывается). В обыденных рассуждениях он даже не указывается, а подразумевается интуитивно. Но такого рода аналогии часто применяются в обыденных рассуждениях.

Другой пример. Обычно рассматривают укороченную фразу «время — деньги» и считается, что она подтверждает, что время тоже можно тратить, экономить, вложить. Но это ограниченная метафора, потому что время нельзя заработать, поделить на проценты. И это как раз свидетельствует о том, как метафора на самом деле сформирована. Утверждением-источником метафоры является, скорее, утверждение «деньги дает работа» или аналогичное. Далее следует предикат установления связи «работа требует времени». Отсюда возникает метафора «время можно тратить как деньги» и т.д. Но так как связь времени и денег установлена через работу, то время нельзя заработать (получается тавтология) или получить проценты (с работы проценты не платят, в отличие от денег), поделить временем тоже не получится, потому что нам платят за нашу работу, а не за выполненную кем-то другим. И таким образом приведенная схема вывода метафоры обосновывает также и ограничение на применение метафоры.

Из известных теорий к предлагаемой гипотезе формирования метафоры наиболее близка теория концептуального смешения Жюль Фоконье [11]. Она исследует, как можно смешивать различные концепции для создания новых значений, подчеркивая роль аналогии в этом когнитивном процессе. Эта идея развивается в теории метафоры Ж. Фоконье, в которой такое смешение называется блендом. Но Ж. Фоконье не описывает конкретный механизм, способствующий осуществлению смешения и образованию метафоры через аналогию, который бы позволил его реализовать, что вполне осуществимо в предлагаемой методе [1].

Многие ученые отмечают связь аналогии и метафоры, отмечая, что сама форма метафоры часто выглядит как указание на аналогию, например, «жизнь как путь». Но никто не приводит конкретного способа формирования метафоры через аналогию так, чтобы это можно было повторить в программе и тем самым экспе-

риментально подтвердить гипотезу. Ниже будет показано, как это можно сделать, используя метод из предыдущей нашей статьи [1], и в данной интерпретации формирования метафоры это становится возможным повторить в программе.

Онтология

Онтология построена на обобщении, когда много разных объектов помещаются нами в один класс, а этот класс вместе с другими похожими — в класс более высокого уровня общности. В общем смысле это позволяет нам облегчить передачу информации, когда можно сказать сразу о предикате к множеству объектов одной фразой и сэкономить память. Как же производится обобщение в онтологии? Конечно же, по общим признакам и отношениям входящих в класс объектов. Но любой признак или отношение — это предикат, выраженный прилагательным или глаголом. И те понятия, что включены в общий класс, являются аналогами по ним. Это могут быть отдельные отношения для всех включаемых объектов или множество отношений и признаков, чуть различающихся для разных объектов, как, например, в классе «минералы». В принципе любое понятие, выражаемое лексемой, является, как известно, каким-то классом, кроме персональных имен, например, пароход «Иван Сverdлов».

Схожесть, аналогичность объектов, входящих в один класс, мы чувствуем интуитивно и стремимся назвать. Как же это происходит? Это ключевой вопрос образования онтологий, как не только выявления аналогии между двумя словами, а образования класса из них как нового объекта, объединяющего их общие признаки. Это важный процесс именно речевого мышления, которое позволяет нам образовать новые понятия и охватить ими больше объектов, чтобы иметь возможность построить рассуждения с ними.

Образование имен классов не случайно. Обычно оно производится по наиболее характерному, распространенному общему предикату объектов класса. В приведенном выше примере про цвет имя класса цветов образовалось как наиболее распространенный предикат к существительному «цветок»: «красный цветок, голубой цветок, желтый цветок». Цвет и стал именем класса окраски цветов, так как более разнообразно и ярко показывал весь спектр цветов. Надо отметить, что само понятие «цветок» уже является обобщением всех луговых растений с яркой окраской, что гово-

рит о том, что одно обобщение (растений) помогает делать другие (цветов). Другим примером могут служить понятия, сформированные по общим предикатам с глаголами, что более распространено. Например, все растения растут, а животные живут. Понятия образованы как субстантивы от общих глаголов, так как называть классы принято существительным. Иногда это бывают более сложные предикаты, как «паровоз» — то, что движется на паровой машине. Это достаточно устойчивый способ называть классы, хотя некоторые случаи требуют исследования этимологии слов, чтобы узнать причину называния класса. Например, транспорт: это слово образовано от старофранцузского *transporter* («перевозить или переносить»), что является общим предикатом для всех транспортных средств.

Дедукция и индукция

Выше была рассмотрена онтология, которая объединяет в классы единичные объекты и называющие их понятия. Индукция же объединяет целые утверждения в общий «закон» или утверждение более общего порядка (содержащее такие кванторы, как «все», и имена классов). Важность аналогии в индукции отмечал еще Р. Карнап [10]. По его мнению, рассуждения по аналогии позволяют нам делать выводы о свойствах или поведении одного объекта на основе его сходства с другим объектом, свойства которого известны. Р. Карнап подчеркивал, что для того, чтобы индуктивные аргументы были применимы, они должны включать в себя аналогии. Это означает, что признание сходства между случаями является основополагающим для формирования обоснованных индуктивных заключений.

Индуктивный вывод по аналогии схож по сути с формированием названия класса, описанного выше. Например, серия утверждений «Дерево тянется к солнцу, трава тянется к солнцу, куст тянется к солнцу» становится общим индуктивным выводом «растения тянутся к солнцу» на основе замены многочисленных имен экземпляров класса на одно наименование, что создает общее индуктивное утверждение. То есть аналогия отдельных понятий утверждений создает обобщение утверждения до индуктивного «закона», который в дальнейшем используется как основа для обратного, дедуктивного вывода от общего утверждения к утверждениям с экземплярами класса.

Именно такая возможность глубокого обобщения по аналогии позволяет создать понятия, недоступные для нейросетей. Например, одна из самых больших проблем беспилотных автомобилей заключается в том, что они не могут понять, что считается препятствием. Если вы видите перевернутый фургон перед своей машиной, вы не собираетесь проезжать сквозь него. Автопилот решит проехать, потому что не знает, что это препятствие вообще, а примера перевернутого фургона не оказалось в его обучающей выборке (реальный случай с автомобилями Tesla). Как в представляемом методе аналогии образуется понятие препятствия на дороге, перед которым надо остановиться? Есть отдельные факты, которые сильно обобщаются: «Дерево препятствует проезду, шлакоблок препятствует проезду, машина препятствует проезду». Так образуется, с одной стороны, понятие «препятствия», являющееся общим классом предметов, которые могут быть на дороге и препятствовать проезду. В этот класс по общим признакам с его экземплярами может попасть любой неподвижный и массивный (общие признаки) предмет.

Надо заметить, что название общего класса не похоже на название экземпляров и содержит гораздо меньше общих предикатов с экземплярами класса, чем экземпляры между собой. То есть понятие класса не аналогично экземплярам. Этот парадокс может ввести в заблуждение, но интуитивно мы знаем, по каким распространенным предикатам образовано общее понятие класса, и именно это позволяет нам связать общее наименование класса с его экземплярами. То есть связь сохраняется от экземпляров к понятию класса, а не наоборот. Мы как бы в дедукции вспоминаем, как образовали класс.

Обучение

Человек может начать распознавать новый класс объектов просто по его описанию. Это свойство называется композициональностью, о нем много пишет сейчас Франсуа Шолле [6]. По его мнению, композициональность недоступна нейросетям. По мнению Шолле, композициональность доступна нейросетям, иначе эти системы могли бы отвечать на комбинированные вопросы, которые не встречались в таком виде в обучающих данных. Но уровень общения у нейросетей в силу их архитектуры не очень глубокий, так что они не могут создать такие абстрактные понятия, как «препят-

ствие». А обучение во многом построено на композициональности, то есть на построении нового описания явления как варианта известного. Например, описания чего-либо в словарях и в обыденных рассуждениях строятся по четкой схеме указания известного понятия, общего с новым (родового) и отличительной черты нового. Это новая композиция. Например, «осел — это лошадь с длинными ушами». Или «водопад — это река, падающая вниз» (обыденное объяснение). Как видно из примера, используется общее понятие для объяснения нового, а мы знаем, что общее понятие мы получаем за счет аналогии, как показано выше (онтологии).

Но не только это конструирование позволяет нам обучаться и открывать новое. Еще более абстрактный уровень, построенный на аналогии, выводит на верхний уровень абстрактные отношения, наиболее общие и применимые для всех понятий и лексем. Это отношения отрицания, положения относительно друг друга чего-либо, родства и т.п. Причем это такие отношения, которые имеют обратимые отношения в одной схеме, то есть прямое отношение подразумевает возможность обратного (не во всех случаях). Например, положение над и под чем-то. Все предметы в комнате находятся одновременно на полу и под потолком. У женщины есть дочь, и женщина — сама дочь какой-то женщины (обратное отношение). Отсюда видно, что такие отношения — это максимальная аналогия, аналогия всех объектов, которые могут быть в комнате, всех женщин, которые кем-то рождены. Это позволяет строить систематические отношения, о которых писали Дж. Фодор и З. Пилишин [7] как об аргументе систематичности, недоступном в коннекционизме. Теперь мы подошли к тому, чтобы объяснить, как это реализуется в языке методами дистрибутивной семантики и аналогии.

Заключение

В описанных выше способах мышления применяется один прием — замена в схеме рассуждения как последовательности предикатов одного из понятий (лексемы) на аналогичное, которое надо объяснить, то есть установить известные из схемы источника отношения между понятиями целевого явления. Заменяемое понятие задает весь концепт нового целевого явления, что позволяет наложить на него схему отношений в источнике. Замена одного из понятий на аналогичное легко реализуема [1]. Это позволяет

сделать предположение о возможности проверки гипотезы через реализацию.

Сам факт того, что приведенные выше отдельные процессы мышления аналитически могут быть описаны при помощи аналогии, позволяет сделать утверждение, что это не разные способы мышления, а разные проявления одного способа работы интеллекта. Мы лишь по-разному их называем при проявлении их разным способом. Ведь даже распознавание — это нахождение аналога той схеме предмета, которая уже известна по ранее полученным примерам. Это говорит о фундаментальном характере такого утверждения, в таком выводе и заключается цель данной статьи.

И третье утверждение, вытекающее из исследования. Для создания аналогии требуется не только структурное сходство хотя бы по одному отношению, но и совпадение прогнозируемого отношением свойства с фактически обнаруживаемым. Это совершенно другой подход «предположение-подтверждение» из опыта, а не просто совпадение по совокупности отношений, что с философской, гносеологической точки зрения имеет принципиальное значение и будет прослежено в дальнейших статьях.

Литература

1. Хомяков А. Б., Чижик П. Новый способ нахождения аналогов как возможность исследования языка, мышления и построения систем искусственного интеллекта // Философские проблемы информационных технологий и киберпространства. 2024. № 1. С. 77-88. <https://doi.org/10.17726/philIT.2024.1.5> (Khomyakov A. B., Chizhik P. A new way of finding analogues as an opportunity to study language, thinking and build artificial intelligence systems // Philosophical Problems of IT & Cyberspace (PhilIT&C). 2024. № 1. P. 77-88. (In Russ.). <https://doi.org/10.17726/philIT.2024.1.5>.)
2. Forbus K. D., Falkenhainer B., Gentner D. The Structure-Mapping Engine: Algorithm and Examples // Artificial Intelligence. Volume 41, Issue 1, November 1989. P. 1-63.
3. Kokinov B., Petrov A. Integration of Memory and Reasoning in Analogy-Making: The AMBR. <https://www.researchgate.net/publication/2593827>.
4. Gentner D., Holyoak K. J., Kokinov B. N. The Analogical Mind Perspectives from Cognitive Science. MIT Press, 2000541 pages.
5. Chollet F. It's Not About Scale, It's About Abstraction. https://youtu.be/s7_NlkBwdj8?si=ENkRbgMZMI84MaJg.
6. Fodor J. A., Pylyshyn Z. W. Connectionism and cognitive architecture: A critical analysis // Cognition, Volume 28, Issues 1-2, March 1988. P. 3-71. <https://www.sciencedirect.com/science/article/abs/pii/>

0010027788900315.

7. Hofstadter D., Mitchell M. The copycat project: A model of mental fluidity and analogy-making. <https://www.researchgate.net/publication/243649607>.
8. Forbus K. D., Gentner D., Law K. MAC/FAC: A model of similarity-based retrieval // *Cognitive Science*, Volume 19, Issue 2, April–June 1995. P. 141-205.
9. Hesse M. *Analogy and Confirmation Theory* // Published online by Cambridge University Press: 14 March 2022, <https://www.cambridge.org/core/journals/philosophy-of-science/article/abs/analogy-and-confirmation-theory>.
10. Fauconnier G., Lakoff G. On Metaphor and Blending // *Cognitive Semiotics*, 5(1-2), December 2009.
11. Hofstadter D. *Fluid Concepts and Creative Analogies* // Harvester Wheatsheaf, 1995. 250 p.
12. Holyoak K. J., Thagard P. The analogical mind // *American Psychologist*, 1997. P. 35-44.

УДК 32.019.5, 004.9

DOI 10.17726/phillIT.2024.2.3



Трансформация политического участия в контексте появления цифровых политических институтов

Стешук Богдан Игоревич,

аспирант 1-го курса,

Вологодский государственный университет

Вологда, Россия

bogdan353590@gmail.com

Аннотация. В статье рассматривается влияние цифровизации на политическое участие через призму новых цифровых политических институтов. Активная цифровизация делает политическое участие более доступным и удобным. Однако его доступность создает эффект «спускового клапана», не позволяя аккумулировать значительное недовольство для выдвижения более существенных политических требований. Затрагивается и проблематика публичности цифровых институтов политического участия. Публичность повышает шанс решить проблему посредством привлечения внимания ответственных лиц, но в то же самое время данный механизм фактически приводит к консервации неэффективных государственных институтов.

Ключевые слова: цифровые политические институты; трансформация политического участия; слактивизм; публичность.

Transformation of political participation in the context of the emergence of digital political institutions

Steshuk Bogdan Igorevich,

first-year post-graduate student,

Vologda State University

Vologda, Russia

bogdan353590@gmail.com

Abstract. The article examines the impact of digitalization on political participation through the lens of new digital political institutions. Active digitalization enhances the accessibility and convenience of political engagement; however, this accessibility creates a «pressure valve» effect, preventing the

accumulation of significant dissatisfaction that could lead to more substantial political demands. The issue of the publicity associated with these digital political institutions is also addressed. Publicity increases the likelihood of addressing issues by attracting the attention of responsible parties, yet it simultaneously tends to preserve ineffective governmental institutions.

Keywords: digital political institutions; transformation of political participation; slacktivism; publicity.

Введение

Активная цифровизация современного общества приводит к появлению новых (цифровых) политических институтов и соответственно к преобразованию социальных практик, в том числе и форм политического участия. Цифровые политические институты, с одной стороны, делают формы политического участия удобнее и доступнее, с другой стороны, их доступность и простота создают эффект «спускового клапана», не позволяя аккумулировать какой-либо значимый уровень неудовлетворенности, который заставляет людей переходить от решения конкретных «ad hoc» проблем к требованиям более общего характера. Противоречивыми социальными эффектами обладает и публичность, свойственная цифровым политическим институтам. Она привлекает внимание к проблеме и затрудняет ее игнорирование со стороны ответственных лиц, однако, вместе с тем, она фактически консервирует неэффективные механизмы реагирования и принятия социально значимых решений.

Цель научного исследования состоит в выявлении того, каким образом появление цифровых политических институтов влияет на политическое участие.

Материалы и методы исследования. Исследование является результатом анализа преимущественно зарубежной литературы, связанной с влиянием цифровизации на политические процессы, а также осмысления практик цифрового политического участия, которые активно используются гражданами и организациями в России.

Одной из важнейших теоретических основ данного исследования послужила книга австро-американского экономиста Альберта фон Хиршмана «Выход, голос и верность. Реакция на упадок

фирм, организаций и государств» [1]. Хиршман предполагает, что у людей существуют две принципиально разные формы реакции на любое раздражающее воздействие со стороны организаций и институтов, в том числе политических: «Выход», означающий выбор аналогов, которые приносят дискомфорт организации или институтам (в случае с политическими институтами речь может идти о переезде), и «Голос», означающий использование человеком системы обратной связи для решения проблемы той или иной организации/института (жалоба, участие в голосовании и т.д.). Сам Альберт Хиршман замечал, что «Выход» является преимущественно «экономическим» решением, в то время как «Голос» представляет собой основной механизм систем обратной связи, в том числе и политической системы. «Выход» позволяет устранить раздражение с минимальными издержками (можно пойти в соседний магазин), «Голос» же предполагает существенные издержки (приходится формулировать жалобу на продавца, разговаривать с руководством и т.д.), однако эта форма реакции служит для организаций и институтов стимулом к изменениям.

В рамках данного исследования мы также попытались проанализировать нижеописанные цифровые политические институты с точки зрения того, какую именно форму политической реакции на раздражение они стимулируют. В первую очередь речь идет о платформах с возможностью создавать онлайн-петиции по наиболее острым вопросам. В качестве примеров можно привести Change.org, Avaaz, Russian Change. Данные платформы активно используются для привлечения внимания к социальным и экологическим проблемам. При этом зачастую данные проблемы носят сугубо локальный и даже личный характер [2]. Например, речь может идти о тяжелом положении отдельной семьи, загрязнении конкретного водоема или вырубке деревьев в месте отдыха. Онлайн-петиции позволяют использовать медийный ресурс для того, чтобы оказать давление на конкретных представителей власти или бизнеса. Однако некоторые исследователи называют такую форму гражданской активности слактивизмом (от английского slack — лентяй) и критикуют ее за поверхностность и попытку подменить реальные политические или социальные преобразования, требующие борьбы, возможностью почувствовать свою причастность к важным изменениям при помощи нажатия пары кнопок.

Такая форма гражданского активизма, не требующая серьез-

ных затрат и несущая минимальные риски, имеет и положительный эффект, давая людям возможность повлиять на что-то конкретное и стать более осведомленными о тех или иных проблемах. Стоит заметить, что существование таких платформ может самым прямым образом влиять и на другие общественные институты. Нередко онлайн-петиции или привлечение СМИ используются в качестве последнего механизма решения проблемы, после того, как все другие механизмы, в частности обращения в профильные органы власти, личные приемы у депутатов и глав администраций, ни к чему не привели.

Публичность — это оружие слабых. Привлечение внимания к проблеме, вероятно, поможет ее решить, и такое решение может оказаться полезным для конкретной группы людей, но крайне пагубным для всего общества в целом. Вместе с решением проблемы уйдет неудовлетворенность и желание каких-либо более значимых перемен. Онлайн-петиции таким образом служат своеобразным клапаном, выпускающим пар социального напряжения, и тем самым консервируют неэффективные институты, так как у людей не остается какой-либо потребности в дальнейших преобразованиях.

С 2020 года в России появилась система Центров управления регионами (ЦУР), цель которых заключается в выстраивании эффективной обратной связи между властью и обществом. ЦУРы позиционируются как инструмент прямой коммуникации жителей и власти с целью решения и предотвращения проблем. Сотрудники таких центров занимаются мониторингом, обработкой и анализом обращений и сообщений жителей, обеспечивают межведомственное взаимодействие органов власти для максимального сокращения времени получения ответа и решения проблемного вопроса гражданина. «Работаем, чтобы сделать государство ближе к людям, создавая экосистему цифрового общения между гражданами страны и органами власти всех уровней», — говорится на сайте АНО «Диалог».

Данная система заменила отдельные региональные инициативы по электронному взаимодействию граждан с органами исполнительной власти. Отличительной чертой Центров управления регионами является их проактивность. Специалисты Центров при помощи специальных алгоритмов на постоянной основе осуществляют мониторинг комментариев, публикаций и иных форм выражения мнения в Интернете, что позволяет выявить проблемы,

которые беспокоят жителей той или иной местности, оперативно связаться с ними и передать информацию о проблеме в соответствующие органы власти [3].

Безусловно, такая платформа повышает эффективность работы органов власти, особенно органов местного самоуправления, однако данная система, так же как и платформы для онлайн-петиций, имеет механизм консервации политических институтов, однако данный механизм более изощрен. Решая наиболее насущные и бытовые проблемы граждан в проактивном режиме, данные центры, во-первых, стремятся упразднить любую мотивацию граждан к самоорганизации и какому-либо коллективному действию, в том числе связанному с выражением недовольства, во-вторых, упраздняют любое напряжение, которое при его накоплении может выйти за границы решения местных проблем и обратиться к куда более фундаментальным вопросам политических свобод, социальной справедливости, распределения благ и т.д. Таким образом, система ЦУРов аккумулирует все социальное напряжение в рамках решения местных проблем, не позволяя этому недовольству переродиться в требования политических перемен.

С 2013 года в России работает портал Regulation.gov, который позволяет всем заинтересованным сторонам принять участие в оценке регулирующего воздействия того или иного проекта нормативного правового акта. Данная платформа дает возможность ознакомиться с текстом важных нормативных правовых актов еще до вступления их в силу и принять участие в оценке качества их проработки и возможных последствий. Regulation.gov активно используется крупным бизнесом для отслеживания возможных регуляторных рисков, а также для продвижения своих интересов путем составления экспертных заключений о том или ином нормативном правовом акте [4].

Благодаря таким системам взаимодействие власти и бизнеса становится прозрачнее, улучшается качество регулирования, а также облегчается возможность участия экспертов в различных областях знаний в формировании различных нормативных правовых актов. Важно заметить, что, отстаивая свои интересы, крупный бизнес также представляет и интересы сотен тысяч людей, занятых в той или иной отрасли [5], т.к. их материальное состояние напрямую зависит от возможности бизнеса нормально функционировать в рамках адекватного нормативного регулирования.

Принципиальная особенность платформы Regulation.gov от других цифровых платформ политического участия заключается в том, что сама ее механика, а также высокая осведомленность обеих сторон (государства и бизнеса) не позволяют подменить реальное политическое участие решением локальных проблем и использовать манипулятивные практики. Анализ данной платформы приводит к выводу о том, что на трансформацию политического участия влияет не столько сам характер взаимодействия акторов, сколько архитектура и механики тех платформ, на которых они взаимодействуют, а также компетентность взаимодействующих сторон.

Regulation.gov является платформой, на которой в обязательном порядке публикуются проекты всех значимых нормативных правовых актов, что не позволяет одной из сторон привлекать больше внимания к одной проблеме в ущерб другой посредством активной публикации одной информации и сокрытия другой. Кроме того, в рамках платформы происходит взаимодействие двух крайне компетентных субъектов: государства и крупного бизнеса (самостоятельно или в лице объединений). Заметим, что крупный бизнес ведет себя достаточно организованно, в России на данный момент активно действует около 300 бизнес-объединений [6], которые активно продвигают интересы своих членов в разных ветвях государственной власти.

Цифровая платформа в данном случае является лишь инструментом, которым бизнес пользуется, опираясь на свою высокую компетенцию, а также на собственные институты выработки и принятия решений, лишь связующим звеном между государством и иными институтами [7]. В этом и заключается принципиальное отличие Regulation.gov от остальных обозначенных в тексте платформ. Regulation.gov не подменяет собой классические институты выработки и принятия решений, а лишь дополняет их, в то же самое время иные платформы имеют дело не с институтами, а с индивидами и фактически подменяют собой существующие институты политического участия.

Столь подробный анализ одной из платформ был необходим для того, чтобы продемонстрировать природу тех негативных последствий, которые цифровые платформы оказывают на политическое участие. Причины этих последствий лежат, в том числе, за пределами цифрового пространства и связаны как с компетентно-

стью самих субъектов, так и с наличием нецифровых институтов выработки и принятия решений.

Мы неслучайно отметили высокую солидарность и способность к кооперации бизнеса. Оба эти свойства помогают ему использовать цифровые платформы как полезный инструмент. В то же время большинство платформ имеют дело с отдельным индивидом (пользователем), компетентность и ресурсы которого (в том числе ресурсы внимания и времени) существенно ограничены. В таких условиях платформы становятся не инструментом взаимодействия одних институтов с другими, а самими институтами, в рамках которых происходит отбор проблематики, выработка общих позиций и принятие решений.

В условиях разрозненности, ограниченности ресурсов и отсутствия/неконкурентоспособности (высокая сложность и длительность, систематическое отсутствие положительных результатов и т.д.) иных институтов политического участия проявляется такое явление, как цифровое доминирование, то есть существенное, иногда преобладающее влияние на принятие решений алгоритмов и механик самой платформы, заложенных (намеренно или нет) в нее разработчиками.

Наиболее концептуально данное явление определила американский философ и социальный психолог Шошана Зубофф. Она назвала современную бизнес-модель цифровых платформ «надзорным капитализмом» [8], описывая ее как систему, в которой человеческий опыт (выраженный в данных о его поведении) становится главным ресурсом, а его накопление — важнейшей задачей всевозможных сервисов. Наиболее явным и вместе с тем хрестоматийным примером стал описанный Ш. Зубофф случай с игрой «Pokemon Go», которая поощряла игроков посещать различные места для поиска цифровых аватаров. За этим весьма безобидным развлечением скрывался коммерческий интерес магазинов и заведений общественного питания. Известно, что владельцы платили разработчикам игры для того, чтобы на территории их заведений цифровые аватары покемонов появлялись чаще, чем в других местах. Это позволило компаниям увеличить число посетителей и потенциальных клиентов. Несмотря на то, что исследовательница практически не касалась вопросов политического участия посредством цифровых платформ, ее выводы относительно модели функционирования цифровых платформ касаются и платформ политического участия.

Платформы, нацеленные на вовлечение в политическое участие отдельных индивидов, строятся на глубокой асимметрии [9] между объемом внимания и компетентности отдельных индивидов и объемом информации, которая генерируется на платформе. В условиях отсутствия альтернативных механизмов отбора проблематики и выработки решений ключевым механизмом данных платформ становится внимание. Проблема, собирающая наибольшее внимание, выигрывает в конкурентной борьбе и получает возможность быть решенной. Очевидно, что наибольшее внимание всегда будут привлекать не отвлеченные и сложные проблемы, связанные с принципиальными политическими вопросами (налогообложением, территориальным устройством и т.д.), а конкретные и насущные проблемы отдельных пользователей.

Стоит упомянуть, что такой механизм отбора проблематик и выработки общих позиций сильно подвержен манипуляционным практикам [10] и позволяет элегантно пренебрегать рассмотрением одних проблем в угоду другим. Описание и анализ этих методов выходят за рамки данной работы, однако само понимание уязвимости таких механизмов позволяет видеть те риски, которые с собой несут цифровые платформы политического участия.

Заключение

Несмотря на простоту и доступность политического участия, которые приносят вместе с собой цифровые политические институты, они способствуют концентрации внимания общества на единичных бытовых проблемах, не позволяя накапливаться должному объему раздражения, которое в своем итоге должно привести к требованиям более высокого порядка. Цифровые политические институты могут способствовать консервации неэффективных средств обратной связи и систем выработки социально-политических решений за счет быстрого решения проблемы и привлечения внимания со стороны вышестоящих инстанций.

Вместе с тем результат функционирования платформ цифрового политического участия принципиально зависит от компетентности субъектов взаимодействия, архитектуры и механики платформ, а также наличия альтернативных механизмов отбора проблематик и выработки общих позиций. В большинстве случаев цифровые платформы не просто дополняют существующие политически институты, а фактически заменяют их. Как новые инсти-

туты, они предлагают новые способы отбора проблематик и выработки общих позиций, которые радикальным образом зависят от распределения человеческого внимания. Подобные механизмы несут в себе риски выхолащивания какого-либо обсуждения сложных и принципиальных политических проблем.

Литература

1. Хиршман А. Выход, голос и верность. 1-е изд. М.: Новое издательство, 2009. 156 с. (Hirschman A. Exit, Voice, and Loyalty. 1st ed. M.: Novoe Izdatel'stvo, 2009. 156 p.)
2. Иваненко Е. А. Change.org как форма «случайной политики»: анализ русскоязычных петиций // Полис. Политические исследования. 2022. № 2. С. 52-57. (Ivanenko E. A. Change.org as a form of «random politics»: analysis of Russian-language petitions // Polis. Politicheskie Issledovaniya. 2022. № 2. P. 52-57.)
3. Чугунов А. В. Развитие системы электронного участия в России в начале 2020-х: роль социальных медиа и центров управления регионами // Политэкс. 2022. № 2. С. 108-136. (Chugunov A. V. Development of the electronic participation system in Russia in the early 2020s: the role of social media and regional management centers // Politex. 2022. № 2. P. 108-136.)
4. Медоева Б. К. Оценка регулирующего воздействия (ОРВ) и оценка фактического воздействия (ОФВ) как институты повышения качества эффективности нормативных-правовых актов // Теоретическая и прикладная юриспруденция. 2023. № 4 (18). С. 52-61. (Medoeva B. K. Assessment of Regulatory Impact (ARI) and Assessment of Actual Impact (AAI) as institutions for improving the quality and effectiveness of regulatory legal acts // Teoreticheskaya i Prikladnaya Yurisprudentsiya. 2023. № 4 (18). P. 52-61.)
5. Павроз А. В. Институт лоббизма в современных демократических обществах // Политическая экспертиза: ПОЛИТЭК. 2014. № 4. С. 158-170. (Pavroz A. V. The Institute of Lobbying in Modern Democratic Societies // Politicheskaya Ekspertiza: POLITEX. 2014. № 4. P. 158-170.)
6. Рейтинг ассоциаций и бизнес объединений // КРОС URL: https://www.cros.ru/ru/exploration/landscape_part_detail/1258 (дата обращения: 04.12.2024.).
7. Негуляев С. В. Цифровизация технополитики совместного управления: кейс платформы «госуслуги. решаем вместе» // Политика и общество. 2022. № 2. С. 37-45. (Negulyaev S. V. Digitalization of Technopolitics in Collaborative Governance: Case of the “Gosuslugi. Reshaem Vmeste” Platform / S. V. Negulyaev // Politika i Obshchestvo. 2022. № 2. P. 37-45.)
8. Zuboff S. The Age of Surveillance Capitalism. New York: Public Affairs, 2019. 704 с.
9. Рыжкова М. В., Чиков М. В. Институциональная природа цифро-

- вых платформ // Вестник Бурятского государственного университета. Экономика и менеджмент. 2019. № 4. С. 72-81. (Ryzhkova M. V., Chikov M. V. Institutional Nature of Digital Platforms // Vestnik Buryatskogo Gosudarstvennogo Universiteta. Ekonomika i Menedzhment. 2019. № 4. P. 72-81.)
10. Лопатин С. В. Политический фрейминг как стратегия манипулирования политическим сознанием // Дискурс-Пи. 2019. № 1 (34). С. 68-77. (Lopatin S. V. Political Framing as a Strategy for Manipulating Political Consciousness // Diskurs-Pi. 2019. № 1 (34). P. 68-77.)

УДК 004.9, 81'32

DOI 10.17726/phillIT.2024.2.4



Эволюция методов обработки естественного языка

Беседина Анастасия Юрьевна,

студентка 1-го курса магистратуры

кафедры теории и практики перевода

Института переводоведения, русистики и многоязычия,

Пятигорский государственный университет

Пятигорск, Россия

nastya.besedina.02@mail.ru

Аннотация. Обработка естественного языка (Natural language processing, NLP) претерпела значительные изменения в своей методике, что отражает прогресс в области вычислительных технологий и когнитивных исследований. В данном обзоре мы осветим ключевые моменты эволюции методов анализа естественного языка. В статье затрагивается тема первых разработанных систем NLP, приведены обоснования причин сложности некоторых обрабатываемых текстов и возможной глубины анализа. Кроме того, здесь описаны не только методы NLP до и после GPT-революции, но и современные тенденции и перспективы в сфере обработки естественного языка. Статья позволяет проследить, как менялось представление о тексте на естественном языке в ходе развития компьютерных методов анализа, а также разобраться, чем является текст в зеркале обработки естественного языка, что в действительности является предметом исследования обработки естественного языка и чего нельзя увидеть глазами простого исследователя, не использующего методы NLP.

Ключевые слова: обработка естественного языка; анализ текста; виртуальные ассистенты; трансформеры; искусственный интеллект; нейронные сети.

Evolution of natural language processing methods

Besedina Anastasia Yuryevna,

first-year graduate student,

Department of Theory and Practice of Translation

Institute of Translation Studies, Russian Studies and Multilingualism

Pyatigorsk State University

Pyatigorsk, Russia

nastya.besedina.02@mail.ru

Abstract. Natural language processing (NLP) has undergone significant changes in its methods, reflecting advances in computing technology and cognitive research. This article reviews the key stages of the evolution of natural language processing methods. The article touches on the topic of the first NLP systems developed, provides justification for the reasons for the complexity of some processed texts and the possible depth of analysis. In addition, it describes not only NLP methods before and after the GPT revolution, but also current trends and prospects in the field of natural language processing. The article allows us to trace how the idea of natural language text has changed during the development of computer analysis methods, as well as to understand what text is in the mirror of natural language processing, what is really the subject of natural language processing research and what cannot be seen through the eyes of a simple researcher who does not use NLP methods.

Keywords: natural language processing; text analysis; virtual assistants; transformers; artificial intelligence; neural networks.

Введение

Исследование и обработка естественного языка (Natural language processing, NLP) представляет собой многоаспектную научную дисциплину, связывающую лингвистику, информатику, когнитивные науки и области искусственного интеллекта. Специалисты этой сферы занимаются созданием вычислительных систем, которые могут осмысливать, анализировать и производить тексты на человеческом языке. Эволюция методов NLP связана с развитием вычислительных мощностей, сменой парадигм и появлением новых теоретических подходов в области когнитивных наук и анализа данных. Данная статья представляет исторический обзор эволюции методов NLP, выделяя ключевые этапы и изменения восприятия явления «текст».

Можно сказать, что обработка естественного языка — это изучение вычислительных методов для автоматического анализа текста и извлечения значимой информации для последующего анализа. Сложность анализа текста «заключается в том, что текст эллиптичен, неполон и насквозь пронизан умолчаниями» [1, с. 106]. **Актуальность** темы статьи обусловлена высоким ростом значимости обработки естественного языка в различных областях деятельности человека. Изучение эволюции методов NLP способно помочь разобраться в работе этой системы, начиная с ее

истоков, и дать достаточно полное представление о развитии методов в историческом срезе.

Цель данной статьи состоит в рассмотрении эволюции методов NLP с исторической точки зрения. **Задача** — проследить, как менялось представление о тексте на естественном языке в ходе развития компьютерных методов анализа.

«Обработка естественного языка — это наука о проектировании методов и алгоритмов, которые принимают или порождают неструктурированные данные естественного языка» [2, с. 19]. Эта сфера знаний занимается взаимодействием между компьютерами и человеческим языком, где главная задача заключается в обеспечении компьютерам способности понимать, анализировать и воспроизводить человеческий язык так же, как это делает человек. В связи с этим, сюда входят такие задачи, как понимание значения (способность извлекать значение из текста, речи или других форм человеческого языка), анализ структуры (распознавание грамматической структуры и синтаксиса языка, включая части речи и построение предложений), создание языка, подобного человеческому (создание естественного, связного и грамматически правильного текста или речи).

Н. Н. Леонтьева описывает различные виды структур, необходимых для понимания текста: лингвистические структуры в предложениях текста, что можно назвать локальным пониманием; семантические сети в целом тексте, что представляет глобальное размытое понимание; информационные структуры в тексте в целом, что дает глобальное обобщенное понимание; и структуры баз данных и знаний, что обеспечивает выборочное специальное понимание [3]. NLP нацелено на преодоление разрыва между человеческим общением и машинным пониманием, способствуя беспрепятственному взаимодействию между людьми и технологиями. При этом, «помимо проблем, связанных с обработкой неоднозначных и вариативных входных данных в системе с плохо определенными и отсутствующими наборами правил», у естественного языка «есть и дополнительные свойства, которые еще больше затрудняют разработку вычислительных подходов на основе машинного обучения: дискретность, композиционность и разреженность» [2, с. 19].

Эволюция методов обработки естественного языка в историческом контексте тесно связана с этапами зарождения и прогресса

машинного перевода, распознавания речи и искусственного интеллекта.

В XVII веке, который ознаменован вкладом ведущих философов, зародился путь развития машинного перевода. Рене Декарт предложил в 1629 году универсальный язык, который содержал эквивалентные идеи на разных языках, однако использующих один символ. В 30-е годы прошлого века были выданы первые патенты на устройства для перевода текстов. Стоит отметить автоматический двуязычный словарь Жоржа Арцруни, который функционировал с помощью бумажной ленты. Питер Троянский предложил концепцию, включающую в себя двуязычный словарь и метод определения грамматических функций между языками. Он основывался на системе языка эсперанто. В 1950 году Алан Тьюринг в своей знаменитой работе «Вычислительная техника и интеллект» представил «тест Тьюринга» как мерило разумности. Этот критерий оценивает способность компьютерной программы на достаточно высоком уровне имитировать человека в процессе текстового разговора с человеком-судьей, настолько, чтобы последний не мог однозначно определить, с кем он общается — с программой или с человеком. Семь лет спустя труд Ноама Хомского «Синтаксические структуры» потряс лингвистику, введя концепцию «универсальной грамматики», представляющую собой систему синтаксических правил. Все вышеперечисленное стало предпосылкой к дальнейшему возникновению новых и изменению существующих на тот момент методов обработки естественного языка.

Первый этап развития методов обработки естественного языка (с 1950-х гг. по 1970-е гг.)

Самые первые разработки систем обработки естественного языка основывались на символьном подходе. Важно понимать, что в то время текст рассматривался скорее как последовательность символов, которую можно анализировать посредством определенных правил и формальных грамматик. Именно тогда стали разрабатываться такие формальные языки или системы описания синтаксиса, как, например, Форма Бэкуса — Наура (БНФ) (Backus-Naur Form — BNF), которые описывали синтаксис языка и анализировали грамматическую структуру предложений, в процессе чего одни синтаксические категории последовательно определялись через другие. Сложность обрабатываемых текстов и глубину ана-

лиза, прежде всего, ограничивало то, что вычислительные мощности были низкими. Чтобы иметь полное представление о развитии обработки естественного языка, необходимо принять во внимание систему ELIZA, которая стала важнейшим достижением того периода, однако, в действительности, демонстрировала ограниченное понимание языка и опиралась на ключевые слова, заготовки, шаблоны, а не на истинное понимание семантики. Ограниченность вычислительных мощностей не давала возможность обрабатывать большие объемы текста и учитывать сложные лингвистические явления, поскольку внимание в большей степени уделялось синтаксическому анализу, а семантический анализ был очень простым. Кроме того, к подходам, основанным на правилах и символьной обработке информации, относятся системы на основе древовидных структур, регулярных выражений и т.д. Специалисты отмечают: «Примерно с 1960 по 1985 год в большинстве областей лингвистики, психологии, искусственного интеллекта и обработки естественного языка полностью доминировал рационалистический подход. Рационалистический подход характеризуется верой в то, что значительная часть знаний, хранящихся в человеческом сознании, не извлекается органами чувств, а фиксируется заранее, предположительно путем генетического наследования» [4, с. 4]. Можно сделать вывод, что сам текст в этой парадигме — последовательность символов, поддающаяся формализации и разбору по заранее заданным правилам. В таком случае «невидимым» оставалось тонкое понимание контекста, многозначности слов и нюансов человеческой речи.

Второй этап развития методов обработки естественного языка (с 1980-х гг. по 1990-е гг.)

По мере роста вычислительных мощностей и накопления больших корпусов текстовых данных стало возможным использование статистических методов. Вместо строго определенных правил, статистические модели теперь обучались на корпусах текстов, выявляя вероятностные связи между словами и фразами. В мире анализа последовательностей возникли новые технологии — скрытые марковские модели (Hidden Markov Model — HMM) и методы n-грамм. Они способны предсказывать следующее слово в предложении, опираясь на предыдущие слова. N-грамма представляет собой последовательность из n элементов, которая может быть

звуковой, слоговой, словесной или буквенной. Такие методы, как n-граммы, скрытые марковские модели и статистический машинный перевод, открыли новые возможности для анализа текстовых данных и помогли усовершенствовать точность анализа текста и решение более сложных задач, как, например, разбор предложений и частотный анализ. Однако статистические модели зачастую не могли в достаточной мере обработать контекст и семантические отношения между словами, что приводило к ошибкам в анализе многозначных слов и сложных предложений. Закономерно, что текст стал рассматриваться как вероятностное распределение слов и фраз, где важность отдельных элементов определяется их частотой и соседством. Здесь соответственно «невидимым» оставались более глубокие семантические отношения, неявные связи между словами и сложные лингвистические явления.

Третий этап развития методов обработки естественного языка (с 2000-х гг. по настоящее время)

Развитие глубокого обучения стало настоящим революционным прорывом в обработке естественного языка. Основой большинства современных систем NLP являются нейросети и глубокое обучение, поскольку данные технологии дают возможность построить модель сложных зависимостей данных и извлечь скрытые закономерности. Благодаря нейронным сетям, особенно рекуррентным нейронным сетям (Recurrent Neural Network — RNN), механизмам долгосрочной краткосрочной памяти (Long Short Term Memory — LSTM), а также трансформерам, стало возможным осуществлять более глубокий анализ и изменение текста. Обратимся к мнению современных исследователей: «Рекуррентные сети являются весьма выразительными моделями для последовательностей и являются, пожалуй, самым полезным, что могут предложить нейронные сети обработке языков. Они позволяют отказаться от марковского предположения, преобладавшего в NLP в течение нескольких десятилетий, и проектировать модели, в которых условиями могут быть целые предложения. При этом они могут при необходимости учитывать порядок слов и не подвержены проблемам статистического оценивания, проистекающим из разреженности данных. Эта возможность дает заметный выигрыш в языковом моделировании — задаче о предсказании вероятности следующего слова в последовательности (или, что то же самое, вероятности

предложения), — которое является краеугольным камнем многих приложений NLP» [2, с. 22]. Векторные представления слов (word embeddings) позволили кодировать семантическую информацию в многомерном пространстве, где схожие по значению слова располагаются близко друг к другу. Механизм внимания в трансформерах позволил моделям фокусироваться на важных частях текста, учитывая контекст и взаимосвязи между различными его частями. Предварительно обученные модели, такие как BERT (Bidirectional encoder representations from transformers), XLNet (авторегрессионная языковая модель, которая выдает на выходе вероятность совместной встречаемости последовательности токенов) и GPT (Generative Pre-trained Transformer), достигли невероятных успехов в различных задачах обработки естественного языка от машинного перевода до генерации текста. Несмотря на то, что нейросети используют большие вычислительные ресурсы и огромные объемы тренировочных данных, проблема смещения и субъективности в обучающих данных также остается актуальной. На текущем этапе развития методов обработки естественного языка текст рассматривается как многомерное семантическое пространство, где связи между его элементами закодированы в сложной структуре нейронной сети. В тексте «невидимым» до сих пор еще остается полное понимание человеческого языка с его нюансами, амбивалентностью и неявным значением.

Рассматривая современные методы обработки естественного языка, важно иметь представление об основных направлениях применения искусственного интеллекта (ИИ) в NLP как важнейшего результата GPT-революции. Одним из ключевых направлений использования искусственного интеллекта в сфере обработки текстовых данных (NLP) является автоматический перевод текстов. В наше время активно применяются переводческие системы на базе искусственного интеллекта, такие как Google Translate и DeepL, в которых для осуществления перевода используются нейросетевые технологии. Эти системы обладают способностью учитывать контекст и семантику слов в тексте, что значительно повышает качество перевода по сравнению с более примитивными статистическими методами. Более того, вспомогательные программы Google Assistant, Siri и Alexa также применяют искусственный интеллект для того, чтобы корректно распознавать и обрабатывать команды пользователей, выраженные на естественном

языке. Эти системы способны отвечать на вопросы пользователей, помогать в организации встреч и заказов, а также поддерживать диалог. Другой важной особенностью является то, что при анализе текстов в различных приложениях стали активно использоваться нейросети. В качестве примера можно привести систему, способную самостоятельно распределять тексты по положительной или отрицательной категории мнений, тональности на основе анализа используемых слов и фраз. Эта технология находит применение, в том числе, в исследованиях рынка и для осуществления оперативного контроля обратной связи от клиентов и динамики активности в социальных сетях. Искусственный интеллект также активно применяется для разделения текстов на группы, выделения ключевых слов и понятий, а также для семантического анализа. Важным направлением использования ИИ в области обработки естественного языка является распознавание речи. Сервисы Google и Speech-to-Text используют нейронные сети для преобразования звука и аудиофайлов в текст. Кроме того, существуют алгоритмы, способные автоматически создавать краткие выжимки даже для очень длинных текстов, выделяя ключевую информацию и сохраняя логическую связь и смысл исходного содержимого. «Нейронные сети дают эффективный механизм обучения, чрезвычайно привлекательный для использования в задачах обработки естественного языка, — утверждают специалисты. — Главный компонент языковой нейронной сети — слой погружения, т.е. отображение дискретных символов на непрерывные векторы в пространстве сравнительно небольшой размерности. В результате погружения слова преобразуются из изолированных дискретных символов в математические объекты, над которыми можно производить различные действия» [2, с. 21].

При обработке текста на естественном языке исследователь обладает всеми необходимыми инструментами и алгоритмами для классификации текста, извлечения информации, проведения тематического моделирования и даже анализа настроений.

Обратимся к методам обработки естественного языка для более детального рассмотрения. Когда возникает необходимость провести классификацию текста, он распределяется по заранее определенным категориям в зависимости от их тематики и наполнения. Следовательно, модели классификации на базе методов обработки естественного языка имеют возможность автоматически распреде-

лять и группировать документы относительно их тематики и тона. С помощью извлечения информации исследователь, как правило, приобретает информацию, обладающую четкой структурой, пусть до этого она и являла собой беспорядочные данные. Распознавание и извлечение определенных событий или отношений способствуют эффективному использованию поиска информации и построению граф-схем знаний. Автоматическое определение тем на основе текстовых данных и распределение документов по темам позволяют получать сгруппированные данные и анализ тенденций. Согласимся с мнением, что «количество текстов на естественном языке, доступных в электронном виде, поистине ошеломляет и увеличивается с каждым днем. Однако сложность естественного языка может сильно затруднить доступ к информации, содержащейся в этом тексте. Современный уровень развития НЛП все еще далек от того, чтобы создавать универсальные представления смысла из неограниченного текста» [5, с. 261].

Рассматривая современные тенденции в области обработки естественного языка и ее перспективы, следует отметить, что сейчас большинство исследований в данной сфере нацелено на многоязычное NLP. К нему относится разработка моделей, способных работать с разными языками. Также существует необходимость в разработке подходов к обработке информации на языках, для которых имеется недостаточно данных. Это включает в себя создание алгоритмов для анализа процессов принятия решений в сложных нейронных сетях. Важны также интеграция NLP с различными направлениями искусственного интеллекта, объединение NLP с компьютерным зрением и робототехникой для формирования более совершенных и умных систем.

Отвечая на вопрос, чего нельзя увидеть глазами простого исследователя, становится понятно, что простой анализ текста, не использующий методы обработки естественного языка, не способен выявить множество взаимосвязей, скрытых от человека при поверхностном анализе, среди которых тонкие смысловые связи между словами, не явные из синтаксиса, зависимость значения слов от контекста и окружающих слов, неявный смысл, передаваемый не только словами, но и интонацией, жестами, и различные значения одного и того же слова в разных контекстах и эмоциональная окраска. Для полноценной работы система анализа текста «должна иметь возможность проанализировать текст, поданный пользовате-

лем на вход, с точки зрения синтаксиса (структуры предложений), семантики (понятий, применяемых в тексте) и прагматики (правильности употребления понятий и целей их употребления). Далее система должна сгенерировать свой отклик во внутреннем представлении, пригодном для логического вывода, и просинтезировать свой отклик на естественном языке» [1, с. 106].

Таким образом, можно сделать вывод, что в контексте развития технологий представления о тексте постоянно менялись. С каждым новым этапом развития методов обработки естественного языка люди все больше начинали интересоваться своим языком и, как следствие, текстом, его непрерывными изменениями и пытались проводить сравнения в историческом срезе. «Естественный язык находится в постоянном движении, адаптируясь к меняющемуся миру, создавая названия и слова для обозначения новых вещей, новых людей и новых концепций» [4, с. 309], следовательно, текст в зеркале обработки естественного языка предстает не просто набором слов, а сложной структурой, отражающей семантику, синтаксис, прагматику и контекст. Именно скрытые взаимосвязи между этими элементами, недоступные простому человеческому наблюдению, являются предметом исследования NLP.

Заключение

1. При первых попытках разработки систем NLP, с 1950-х по 1970-е гг., методы основывались преимущественно на символьном подходе, а текст представлялся последовательностью символов, подлежащей анализу с помощью формальных грамматик и правил. Однако все еще предстояло искать пути к тонкому пониманию контекста, многозначности слов и нюансов человеческой речи.

2. На втором этапе развития методов обработки естественного языка, с 1980-х по 1990-е гг., текст стал рассматриваться как вероятностное распределение слов и фраз, где важность отдельных элементов определялась их частотой и соседством. Глубокие семантические отношения, неявные связи между словами и сложные лингвистические явления сохраняли свою научную непостижимость.

3. Третий этап развития методов обработки естественного языка, с 2000-х гг. по настоящее время, отличается тем, что текст трактуется как многомерное семантическое пространство, в котором связи между элементами закодированы в сложной структуре

нейронной сети. Однако результат оставляет задачу полного понимания человеческого языка с его нюансами, амбивалентностью и неявным значением для будущих исследований.

4. Главной движущей силой существующих на данный момент систем обработки естественного языка можно назвать нейросети, трансформеры и глубокое обучение. К преимуществам таких систем относятся улучшение качества и точности обработки и анализа, автоматизация рутины, персонализация, к вызовам — проблемы с интерпретацией и объяснимостью, требования к данным и ресурсам, этические и социальные аспекты.

5. Анализ текста без использования методов NLP не способен выявить большинство скрытых взаимосвязей, среди которых семантические отношения, контекстуальное значение, прагматика и многозначность.

6. В зеркале обработки естественного языка на данный момент текст является сложной структурой, отражающей семантику, синтаксис, прагматику и контекст.

Эволюция NLP показывает, как постепенный рост вычислительных мощностей и развитие новых методов позволяют все глубже проникать в сложную структуру человеческого языка. Будущие прорывы в NLP будут определяться развитием новых алгоритмов и увеличением доступности вычислительных ресурсов и больших языковых моделей. Данная статья дает полноценную картину этапов эволюции обработки естественного языка в своих методах и позволяет проследить взаимосвязь развития методов анализа и представления о тексте на естественном языке.

Литература

1. Большакова Е. И., Клышинский Э. С., Ландэ Д. В., Носков А. А., Пескова О. В., Ягунова Е. В. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие. М.: МИЭМ, 2011. 272 с. (Bolshakova E. I., Klyshinsky E. S., Lande D. V., Noskov A. A., Peskova O. V., Yagunova E. V. Automatic text processing in natural language and computational linguistics: textbook. the manual. M.: MIEM, 2011. 272 p.)
2. Гольдберг Й. Нейросетевые методы в обработке естественного языка. Пер. с англ. А. А. Слинкина. М.: ДМК Пресс, 2019. 282 с. (Goldberg J. Neural network methods in natural language processing. Translated from English by A. A. Slinkin. M.: DMK Press, 2019. 282 p.)
3. Леонтьева Н. Н. Автоматическое понимание текстов: системы, модели, ресурсы: учеб. пособие. М.: ИЦ «Академия», 2006. 304 с. (Leontieva N. N. Automatic understanding of texts: systems, models,

- resources: a textbook. M.: IC “Academia”, 2006. 304 p.)
4. Manning C. D., Schütze H. Foundations of Statistical Natural Language Processing, 8, [print.]-е изд. Cambridge, Massachusetts: The MIT Press, 2005. 720 p.
 5. Bird S., Klein E., Loper E. Natural language processing with Python. 1st ed-е изд. Beijing; Cambridge [Mass.]: O'Reilly, 2009. 479 с.
 6. Тюрин Д. А., Пальмов С. В. Применение нейронных сетей в обработке естественного языка // Журнал прикладных исследований. 2023. № 7. С. 158-162. (Tyurina D. A., Palmov S. V. Application of neural networks in natural language processing // Journal of Applied Research. 2023. № 7. P. 158-162.)
 7. Jackson P., Moulinier I. Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization // Computational Linguistics. 2003. № 29. P. 510-511.
 8. Jelinek F. Statistical methods for speech recognition. Cambridge (GB): MIT press, 1998. 305 p.
 9. Jurafsky D., Martin J. H. Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 2023. 636 p.
 10. Kumar L. A., Renukay D. K. Deep learning approach for natural language processing, speech, and computer vision: techniques and use cases. First edition-е изд., Boca Raton, FL: CRC Press, 2023. 246 p.
 11. Garg M., Kumar S. Natural language processing and information retrieval: principles and applications. London, New York: CRC Press, 2024. 271 p.
 12. Manning C. D., Raghavan P., Schütze H. Introduction to information retrieval. Cambridge: Cambridge university press, 2008. 528 p.
 13. Everaert M. B. H., Huybregts M. A. C., Chomsky N., Berwick R. C., Bolhuis J. J. Structures, not strings: Linguistics as part of the cognitive sciences // Trends in Cognitive Sciences. 2015. № 19(12). P. 729-743.
 14. Mihalcea R., Radev D. Graph-based natural language processing and information retrieval. Cambridge; New York: Cambridge University Press, 2011. 192 p.
 15. Rothman D. Transformers for natural language processing and computer vision: explore generative AI and large language models with Hugging Face, ChatGPT, GPT-4V, and DALL-E3. Third edition, revised publication Sept 2024-е изд., Birmingham, UK Mumbai: Packt Publishing Ltd, 2024. 693 p.

УДК 82-312.9

DOI 10.17726/phillT.2024.2.5



**Специфика отображения социальных сетей
и компьютерных технологий в фантастической
вселенной Стивена Кинга в начале 70-х гг. XX в. —
начале 20-х гг. XXI в. в исторической ретроспективе**

Каспарян Константин Викторович,

*кандидат исторических наук, доцент, профессор кафедры
исторических и социально-философских дисциплин,
востоковедения и теологии,
Пятигорский государственный университет
Пятигорск, Россия
kasparyan@pgu.ru*

Рутковская Марина Валерьевна,

*кандидат философских наук, доцент, доцент кафедры
исторических и социально-философских дисциплин,
востоковедения и теологии,
Пятигорский государственный университет
Ессентуки, Россия
rutkovskayam@list.ru*

Колесников Илья Николаевич,

*кандидат исторических наук, доцент, доцент кафедры
исторических и социально-философских дисциплин,
востоковедения и теологии,
Пятигорский государственный университет
Ессентуки, Россия
in_kolesnikov@mail.ru*

Аннотация. Данная статья посвящена анализу характерных особенностей процесса отображения виртуальных сетевых площадок и кибернетических технологий в начале 1970-х — начале 2020-х гг. в фантастической вселенной (литературной и кинематографической), созданной С. Кингом — одним из ведущих американских писателей-фантастов современности. Авторы приводят обоснование актуальности и научной новизны изучаемой проблемы. В работе проанализированы специфиче-

ческие черты воздействия электронных коммуникационных сообществ и компьютерных технологий на современный социум; приводится аргументация потребности в осмыслении освещения данной общественной проблемы в фантастическом искусстве — конкретно в литературе и кинематографии — как в одном из дополнительных аспектов прогресса прогностической функции науки; объясняется определение фантастической литературно-кинематографической вселенной С. Кинга в качестве объекта исследования, с учетом ее приоритетного положения в фантастическом жанре в целом. Обозначенная проблема избрана авторами для анализа также с целью опровержения стереотипов, согласно которым произведения С. Кинга и созданные на их основе киноадаптации являются комплексом материалов массовой культуры развлекательного характера, не имеющих какой-либо смысловой значимости. В то же время авторы исследуют кардинальные отличия в освещении изучаемой проблемы в оригинальных литературных трудах С. Кинга и их экранизациях. Данный подход, помимо всего прочего, позволил авторам выявить причины негативного отношения писателя к воплощению его книг на киноэкране. В работе осмыслена специфика процесса трансформации отображения социальных сетей и высоких технологий в фантастической вселенной С. Кинга в течение 1970-х — начала 2020-х гг. в контексте реакции автора на объективные изменения в научно-техническом прогрессе и представлениях о нем в общественном мнении, а также роста их значимости в социально-экономической, духовной и политической сферах жизни человеческого сообщества. На примере фантастической вселенной С. Кинга проанализированы характерные черты эволюции положения экспертов в сфере компьютерных технологий в течение последних пяти десятилетий в контексте их восприятия в литературе и кинематографе. Исследованы моральные критерии использования социальных сетей и кибернетических технологий в общественной жизни и их влияние на развитие современных семейных отношений в восприятии С. Кинга — одного из субъектов формирования общественных отношений в западном социуме.

Ключевые слова: фантастическая вселенная Стивена Кинга; постиндустриальная эпоха; искусственный интеллект; социальные сети; интернетофобия; компьютерофобия; технический прогресс; кибернетические атаки; хакер; авторское право.

**The specifics of displaying social networks and
computer technologies in the fantasy universe of
Stephen King in the early 70s of the XX century — early
20s of the XXI century in historical retrospect**

Kasparyan Konstantin Viktorovich,

*PhD (History), associate professor,
Department of Historical and Socio-Philosophical Disciplines,
Oriental Studies and Theology,
Pyatigorsk State University
Pyatigorsk, Russia*

kasparyan@pgu.ru

Rutkovskaya Marina Valeryevna,

*PhD (Philosophy), associate professor,
Department of Historical, Socio-Philosophical Disciplines,
Oriental Studies and Theology,
Pyatigorsk State University
Essentuki, Russia*

rutkovskayam@list.ru

Kolesnikov Ilya Nikolaevich,

*PhD (History), associate professor,
Department of Historical and Socio-Philosophical Disciplines,
Oriental Studies and Theology,
Pyatigorsk State University
Pyatigorsk, Russia*

in_kolesnikov@mail.ru

Abstract. This article is devoted to the analysis of the characteristic features of the process of displaying virtual network platforms and cybernetic technologies in the early 1970s — early 2020s in the fantastic universe (literary and cinematic) created by S. King, one of the leading American science fiction writers of our time. In this study, the authors provide a justification for the relevance and scientific novelty of the problem under study. This paper analyzes the specific features of the impact of electronic communication communities and computer technologies on modern society. This article provides an argument for the need to comprehend the coverage of this social problem in fantastic art — specifically in literature and cinematography — as one of the

additional aspects of the progress of the predictive function of science. This study provides a reasoned explanation of the definition of the fantastic literary and cinematic universe of S. King as an object of research, taking into account its priority position in the fantasy genre as a whole. This problem was chosen by the authors for analysis in order to reasonably refute stereotypes, according to which the works of S. King and the film adaptations created on their basis are a complex of mass culture materials of an entertainment nature that do not have any semantic significance. At the same time, the authors explore the cardinal differences in the coverage of the studied problem in the original literary works of S. King and their film adaptations. This approach, among other things, allowed the authors to identify the reasons for the negative attitude of the writer to the embodiment of his books on the cinema screen. In this work, the specifics of the process of transformation of the display of social networks and high technologies in the fantastic universe of S. King during the 1970s and early 2020s are comprehended in the context of the author's reaction to objective changes in scientific and technological progress and ideas about it in public opinion, as well as the growth of their importance in socio-economic, spiritual and the political spheres of life of the human community. In this study, using the example of the fantastic universe of S. King, the characteristic features of the evolution of the position of experts in the field of computer technology over the past five decades in the context of their perception in literature and cinema are also analyzed. The work also examines the moral criteria for the use of social networks and cybernetic technologies in public life and their impact on the development of modern family relations in the perception of S. King, one of the subjects of the formation of public relations in Western society.

Keywords: Stephen King's fantasy universe; post-industrial era; artificial intelligence; social networks; internet phobia; computer phobia; technical progress; cyber-attacks; hacker; copyright.

Введение

На рубеже XX-XXI столетий человеческий социум, согласно общепризнанному цивилизационному подходу, перешел к постиндустриальной эпохе — периоду автоматизации и компьютеризации производства, дистанционного обучения и выполнения различных видов работ, онлайн-банкинга, виртуальных сетевых площадок, ориентированных на заполнение досуга, и т.д. По существу, в современных условиях виртуальное пространство и кибернетические технологии приобретают ключевую роль в развитии всех основных сфер общественной жизни — духовной, политиче-

ской, экономической и социальной. Активное применение данных феноменов постиндустриального развития в существенной мере упростило практическую деятельность хозяйственных субъектов, политических структур, научных сообществ, формальных и неформальных социальных групп.

Внедрение компьютерных технологий и созданных с их помощью онлайн-платформ в общественные отношения во многом обусловило формирование мировой экономики как противоречивого комплекса национальных хозяйств и ускорило процесс создания транснациональных корпораций. Кроме того, использование кибернетических инноваций в экономике в значительной мере повысило эффективность расчетно-кассового обслуживания клиентов в деятельности банков, выполнение работы в удаленном режиме, потребительскую активность физических лиц на рынке товаров и услуг.

Результаты анализа современной политической обстановки свидетельствуют, что использование кибернетических новинок приобретает все большее значение в военной сфере, где беспилотные летательные аппараты и иные технические устройства, управляемые дистанционно (совсем недавно воспринимавшиеся лишь в качестве атрибута фантастических произведений), становятся неотъемлемой частью вооружения армий ведущих мировых держав и их младших партнеров. Кроме того, компьютерное планирование является значимым компонентом военных стратегий различных государств, создающих, помимо всего прочего, особые подразделения в своих вооруженных силах с целью противодействия кибернетическим атакам противников.

Интенсификация деятельности виртуального пространства обусловила создание дистанционной формы образовательных услуг. Данный факт обеспечил процесс гуманизации образования и предоставил возможность получения знаний лицам с ограниченными физическими возможностями.

Коммуникационные сетевые ресурсы дают людям возможность личного общения вне зависимости от местонахождения собеседника в любое время суток и создания масштабных сообществ по интересам. Кроме того, естественным предметом обихода многих семей стало использование «умных колонок» — портативных баз данных, способных вступать в диалог со своими хозяевами. Уже не вызывает особого резонанса наличие не только электрон-

ных навигаторов в автомобилях, но и беспилотных пассажирских машин.

Сложно отрицать тот факт, что предпочтение онлайн-играм как основному источнику досуга в современных условиях отдает значительная часть человеческого сообщества вне зависимости от возрастной категории. В то же время базовой аудиторией сетевых игр являются представители молодого поколения. Совершенствование сетевого игрового бизнеса, наряду с ростом значимости социальных сетей в общении индивидов по всему миру, по существу, привело к формированию нового вида человеческой культуры, с собственным сленгом, специфическими правилами поведения, своими вершинами творчества, героями и антигероями. К числу положительных аспектов развития виртуального пространства следует отнести его компенсаторный характер для людей, оказавшихся в одиночестве или испытывающих затруднения при непосредственных контактах. Применение компьютерных технологий в значительной степени расширило возможности кинематографии и мультипликации в формировании качественного визуального ряда.

В то же время виртуальные сетевые площадки и интенсивно прогрессирующие компьютерные технологии, помимо всего вышеперечисленного комплекса положительных качеств, наделены и несколько иными характеристиками, отражающими противоречивую сущность постиндустриальной научно-технической революции. Активное (зачастую в чрезмерной степени) использование социальных сетей в качестве приоритетного и едва ли не единственного способа межличностного и межгруппового неформального общения наряду с зависимостью от использования ресурсов Интернета в поисках необходимых материалов для реализации учебного процесса препятствует развитию логического мышления и навыков долгосрочного усвоения информации. Подобная ситуация не способствует и приобретению навыков общения в реальном режиме, осознанию истинного значения традиционных ценностей. Также следует констатировать, что социальные сети становятся площадкой для лиц, совершающих мошеннические действия, распространяющих экстремистские политические и псевдорелигиозные идеи, подталкивающих людей с неустойчивой психикой (в первую очередь — детей и подростков) к суициду. Наиболее активные интернет-пользователи испытывают склонность к гиподи-

намии, редко проводят время на открытом воздухе, что негативно влияет на их физическое и психическое здоровье.

С учетом приведенных фактов становится очевидным интерес к рассматриваемой проблеме со стороны деятелей искусства (писателей, режиссеров, художников), имеющих отношение к научно-фантастическому жанру, который во многом связан с отображением технологического прогресса. При этом большое число фантастов внесли существенный вклад в популяризацию процесса технического развития, в том числе его кибернетической составляющей, и в проблему осмысления этической стороны использования искусственного интеллекта в общественной жизни с философской или исторической точек зрения. В данном контексте, как представляется, наибольшую значимость имеют произведения писателей Ж. Верна, Г. Уэллса, А. Р. Беляева, Э. Гамильтона, Р. Шекли, Р. Брэдли, Р. Хайнлайна, Ф. Карсака, братьев А. Н. и Б. Н. Стругацких, Ф. Дика, С. Лема, К. Булычева, а также режиссеров С. Кубрика, А. А. Тарковского, Д. Лукаса, Д. Кэмерона, Р. Скотта, братьев Л. и Э. Вачовски.

Как представляется, непосредственное отношение к данной категории деятелей искусства имеет американский писатель-фантаст С. Кинг, который зачастую воспринимается как создатель книг невысокого качества в жанре ужасов, предназначенных для экзальтированных читателей. Однако подобная точка зрения во многом была сформирована на основе изучения не самых качественных экранизаций его романов и повестей (которые сам писатель воспринимает резко негативно) или по причине отсутствия понимания их глубокого смысла, заложенного автором. В реальности многие книги С. Кинга — «Кэрри», «Сияние», «Зеленая миля», «Побег из Шоушенка», «Мизери» и др. — относятся к шедеврам современной художественной литературы с точки зрения как профессиональных критиков, так и читателей; лучшие экранные адаптации его произведений, в которых исполняли роли ключевых персонажей такие выдающиеся актеры, как С. Спейсек, П. Лори, К. Уокен, М. Шин, Д. Николсон, Э. Хопкинс, Т. Хэнкс, Д. Сандерленд, К. Гуджино и др., включены в сокровищницу мирового кинематографа. Подобные результаты обусловлены тем, что С. Кинг рассматривает в своем творчестве глобальные социальные и политические проблемы, а также вопросы экзистенциального характера: проблему конфликта поколений, рост бездуховности и безнравственности

на фоне укоренения стяжательства, религиозный фанатизм и делинквентное поведение индивидов (в первую очередь — маньяков и неадекватных поклонников деятелей искусства), экологические преступления, травлю в социальной группе, творческий кризис писателя.

В наименьшей степени данный автор изучает такие вопросы, как зависимость от социальных сетей и онлайн-игр, технофобия, интернетофобия, компьютерофобия, хакерство, сетевое мошенничество, рост безработицы, связанный с компьютеризацией производства, интернет-троллинг. Некоторые прогнозы С. Кинга относительно развития кибернетики и виртуального характера носят революционный характер; в отдельных случаях они не уступают прогностическим изысканиям Ж. Верна и Г. Уэллса. Кроме того, эволюция воззрений писателя на рассматриваемую научную проблему во многом отражает и многовекторную трансформацию восприятия кибернетического пространства и высоких технологий в общественном сознании на протяжении последних десятилетий. В данном направлении также реализуют свой творческий поиск кинорежиссеры Голливуда, работающие с его литературным материалом, — Б. де Пальма, С. Кубрик, Д. Кроненберг, М. Гаррис, К. Пирс, Д. Э. Келли, Т. Уильямс, М. Флэнаган, К. Уиммер, Л. Бир и др.

Следовательно, с точки зрения авторов данного исследования, осмысление процесса отражения онлайн-площадок и кибернетических технологий в фантастической вселенной С. Кинга обладает не меньшей актуальностью, чем произведенный в предыдущих научных статьях анализ рассматриваемой проблемы в фантастических вселенных «Терминатора» Д. Кэмерона и «Звездных войн» Д. Лукаса.

Цель работы: комплексный анализ отображения роли виртуальных коммуникационных сообществ и компьютерных технологий в литературно-кинематографической вселенной С. Кинга в исторической ретроспективе (начало 1970-х — начало 2020-х гг.).

Материалы и методы исследования: прежде всего, необходимо принять во внимание то обстоятельство, что специфические черты отображения онлайн-пространства и кибернетических устройств в фантастической вселенной С. Кинга рассмотрены, наряду с другими аспектами творчества писателя, в научных работах таких исследователей, как П. А. Горохов, М. М. Давыдова,

Т. А. Дустмурадов, У. Х. Кларк, Э. Маркхэм, Д. Олсон, Д. А. Попов, Л. Роугек, М. Роч, Ф. Солиев, С. Спайгнеси, Н. Ю. Филистова, А. С. Сакова, С. Уайатер, И. В. Шалимова, К. Е. Годунова, М. Хафдал, К. Флоренс. Кроме того, существенную научную ценность в данном контексте представляют мемуары самого автора («Воспоминания о писательском мастерстве»), позволяющие осмыслить непосредственную точку зрения С. Кинга на рассматриваемую проблему.

Следует отметить, что такие специалисты, как Л. В. Бавева, Ю. Л. Варшавская, Е. Гоцило, Б. Гоцило, Е. Ю. Козьмина, А. П. Кошкин, С. А. Мельков, А. А. Манахов, Л. Б. Окавка, К. Э. Разлогов, Д. А. Хрюкин, анализируют в своих исследованиях общие концепции восприятия интернет-площадок и сложных электронно-вычислительных устройств.

Также необходимо принять во внимание тот факт, что ключевые проблемы цифрового пространства и информационных технологий как одного из ключевых компонентов развития современного социума рассмотрены в трудах таких ученых, как А. Н. Афанасьев, Б. Н. Баранов, Э. Берн, Н. С. Овчинникова, Д. Р. Паль.

Результаты исследования и их обсуждение

Прежде всего, следует отметить, что в данном исследовании авторы не ставили перед собой цели проанализировать всю совокупность литературных произведений С. Кинга и их киноадаптаций, с учетом их чрезмерного количества, требующего формата не статьи, а монографического издания. В работе изучены исключительно те произведения фантастической вселенной писателя, с которыми авторы непосредственно ознакомились и, что наиболее важно, которые являются наиболее значимыми единицами данной многоаспектной франшизы.

Как уже отмечалось выше, эксперты рассматривают произведения С. Кинга и основанные на них кинематографические работы прежде всего с точки зрения их влияния на отображение значимых общественных проблем современности и применения фантастического/мистического антуража лишь в качестве своеобразной витрины и зрелищного компонента, позволяющего привлечь к ним внимание представителей молодежи, в существенной степени склонных к освоению книг и фильмов развлекательного характера [16, с. 12].

Кроме того, ученые, исследующие фантастическую вселенную С. Кинга, констатируют, что на протяжении десятилетий как писатель, так и использующие его творчество для создания своих произведений режиссеры четко реагируют на весь комплекс изменений, происходящих в общественном сознании и в сфере научно-технического прогресса, в особенности в ее сегменте, связанном с высокими технологиями и порождаемыми ими положительными и негативными последствиями [10, p. 151].

Таким образом, имеющие место утверждения о примитивности литературных текстов С. Кинга и их кинематографических киноадаптаций, ориентированных на самых невзыскательных в интеллектуальном смысле читателей и зрителей, являются полностью необоснованными и субъективными [6, с. 180].

Следует принять во внимание тот факт, что С. Кинг приступил к занятию профессиональной творческой деятельностью в начале 1970-х годов, в весьма молодом возрасте. При этом будущий писатель создал свои первые любительские рассказы в семилетнем возрасте. Уже в тот период, как отмечают эксперты, юный школьник демонстрировал глубокую заинтересованность в изучении технических новинок. Среди книг, которые он читал, помимо художественных произведений, имелись в наличии многочисленные справочники, включая материалы по электронике [12, с. 1280].

Характерно, что уже его первые произведения получили высокую оценку критиков не только за увлекательный сюжет, но и за привлечение внимания читателей к сложным социальным проблемам, а также за высокий интеллект автора и его глубокие познания — в том числе в вопросах технического развития, проявившиеся, например, в подробном описании существовавших в тот период новейших электронных устройств [5, с. 15].

Первый роман писателя, «Кэрри», был опубликован в 1974 г. Уже в данном произведении проявилась тенденция, характерная для С. Кинга, — жесткая критика религиозного фанатизма и присущей подросткам склонности к организации травли ровесников с отклоняющимся поведением.

Маргарет Уайт, мать главной героини, школьницы Кэрри (Кэриэтты), страдающей от издевательств со стороны одноклассников, испытывает ненависть к окружающему миру, в том числе — к электронным устройствам, которые, по ее мнению, являются порождением антибожественной сущности, так как принцип их ра-

боты не может быть доступен разуму простого человека [4, с. 93]. Позднее С. Кинг отмечал в своих воспоминаниях, что, с его точки зрения, фанатизм на конфессиональной почве, не имеющий ничего общего с реальным вероучением и его постулатами и ритуалами, является, в силу присущего ему невежества, одним из основных противников общественного прогресса и, в особенности, научно-технического. Автор высказывает удовлетворение тем фактом, что режиссер Б. де Пальма, создатель первой киноадаптации «Кэрри», сумел в точности передать данную сторону Маргарет Уайт. Писатель так же высоко оценил мастерство П. Лори, исполнившей роль миссис Уайт (и номинированной на «Оскар» за роль второго плана) [22, р. 30].

С. Кинг акцентирует внимание и на том факте, что после завершения работы над романом расценил его как слабое произведение и выбросил черновик в урну. Ситуация была спасена супругой автора, писательницей Т. Д. Кинг (урожденной Спрус), извлекавшей материал из мусорного ведра и убедившей мужа отправить книгу в издательство. Писатель подчеркивает, что в 1974 г. у него, по понятным причинам, не было электронной копии и, если бы Т. Д. Кинг вернулась домой на несколько часов позже, «Кэрри» не дошла бы до читательской аудитории, а именно успех этой книги послужил основой для его многолетней удачной творческой карьеры.

С. Кинг комментирует жалобы молодых коллег по писательской деятельности на сложности их работы как следствие непонимания преимуществ, которыми они обладают в сравнении с ним в момент начала творчества (в частности, наличие компьютеров и разнообразных электронных баз данных вместо единственной механической печатной машинки) [22, р. 32-33].

Фильм Б. де Пальмы вышел в 1976 г. В 1999 г. на экраны было выпущено продолжение — «Кэрри-2: Ярость», имеющее косвенное отношение к первоисточнику; в 2002 г. была снята вторая экранизация романа. Оба фильма были негативно восприняты критиками и зрительской аудиторией. При этом критические замечания относились к фактору, который неосведомленные специалисты и читатели рассматривают в качестве основного недостатка книг С. Кинга, — максимальный акцент на зрелищности (в данном конкретном случае — активное насыщение кинокартин компьютерными специальными эффектами) при минимальном содержании смыслового компонента [1, с. 85].

В 2013 г. свою кинематографическую версию «Кэрри» выпустила режиссер К. Пирс, которая специализируется на фильмах, отображающих травлю в школе или в отношении лиц, отличающихся от основной части людей особенностями своего поведения, привычек и т.д. Автор данного фильма сумела в не менее убедительной, чем Б. де Пальма, степени показать технофобский характер Маргарет Уайт, чью роль исполнила известная актриса Д. Мур.

При этом, по решению К. Пирс, действие фильма было перенесено из 1970-х в начало 2010-х гг., что позволило ей отразить современные тенденции. Так, в начальной сцене кинокартины, происходящей в душе, издаваемые над Кэрри Уайт школьницы снимают ее на свои смартфоны; позднее неформальный лидер девушек Кристин Харгенсен выкладывает видео в школьном виртуальном паблике и в социальных сетях; в фильме, соответственно, отражена не только травля Кэрри в школе, но и активный интернет-троллинг, который не имеет эффекта лишь по причине того, что Маргарет Уайт категорически запрещает дочери использовать гаджеты даже в учебных целях [9, с. 308].

Позднее, при разбирательстве в кабинете директора школы, именно кадры со смартфона Крис Харгенсен, которые старшеклассница забыла удалить, служат в качестве компромата, позволившего школьной администрации применить в ее отношении формальные негативные санкции, которые в итоге спровоцировали трагедию на выпускном балу [14, с. 281].

Данная кинематографическая работа получила положительный отзыв писателя, в том числе и за смену хронологического периода, в котором происходят события. С. Кинг отметил, что стремление детей и подростков фиксировать издевательства на свои гаджеты и сопутствующий данному занятию интернет-троллинг являются еще более опасной формой травли, способной лишить жизни большое количество людей, чем тот вариант, который он описал в своем дебютном литературном произведении [29, р. 65].

В романе 1978 г. «Противостояние» была описана гибель человеческой цивилизации в результате распространения вируса, описание которого в значительной степени имеет сходство с коронавирусной инфекцией. В самом литературном произведении рассматриваемая тема не была раскрыта, однако «Противостояние» оказало определенное воздействие на кинематографическое искусство и послужило основой для отображения идей режиссе-

ров относительно искусственного интеллекта и социальных сетей [23, р. 102].

В 1994 г. М. Гаррисом была осуществлена первая экранизация в формате мини-сериала. В данной кинематографической адаптации вирус являлся результатом работы над созданием биологического оружия, спроектированным в результате использования новейших компьютерных программ, в расчетах которых была совершена существенная ошибка [28, р. 96].

Как отметил С. Кинг, принявший участие в создании сериала и исполнивший роль второстепенного персонажа Теодора Уэйзека, тот факт, что, согласно сценарию, компьютерный сбой был спровоцирован вмешательством потусторонней силы, не отменяет реалистичности ситуации и аллюзии на опасность всецелого доверия искусственному интеллекту [20, р. 45].

Как представляется, в данном случае следует, прежде всего, провести параллель с творчеством режиссера Д. Кэмерона, создавшего один из самых ярких образов несущего угрозу человечеству электронного разума — системы Скайнет, тем более что писатель одобрительно оценивал данную идею своего соотечественника [15, с. 113].

С. Кинг высказал сожаление, что скромный бюджет сериала не позволил использовать революционные спецэффекты, подобные тем, что были задействованы в «Терминаторе-2» в 1991 г. Характерно, что данный комментарий является едва ли не единственным примером того, когда писатель акцентировал внимание на компьютерной графике. Сам автор вполне аргументированно заметил, что качественные виртуальные эффекты могли в большей степени подчеркнуть как мистический аспект сюжета, так и постапокалиптический характер описанного в нем окружающего мира [18, с. 82].

В 2020 г. режиссер Д. Бун выпустил новый вариант экранизации «Противостояния», адаптированный к современным реалиям (также в виде многосерийного фильма). Сериал вышел на экраны в период активной фазы пандемии COVID-19, по этой причине он был отмечен повышенным интересом со стороны зрительской аудитории и критиков.

При этом, с присущей фантастическому жанру в целом тенденции, данная киноадаптация акцентирует внимание на том обстоятельстве, что коронавирус, при всей его смертоносной сущности, не является предельным вариантом массовой эпидемии. Описыва-

емая в телевизионном фильме 2020 г. пандемия оказалась намного более губительной, чем в реальности. Следует отметить тот факт, что авторы сериала заостряют внимание на неэффективности людей противостоять инфекции даже посредством самоизоляции и работы в дистанционном режиме в случае, когда вирус является продуктом осознанной человеческой деятельности [25, p. 117].

В данном сериале демонстрируется все более возрастающая роль виртуального пространства в общественной жизни. В частности, один из главных героев произведения, Ларри Андервуд (роль исполнил актер Д. Адепо), является известным и успешным музыкантом. Решающее значение в развитии его сольной карьеры имело размещение в Интернете видеоклипа по его авторской песне, в короткий срок набравшего сотни тысяч просмотров, сопровождаемых лайками и положительными комментариями [21, p. 429].

Также в данном сериале раскрывается зависимость людей от инновационных технологий и социальных сетей. Большая часть выживших весьма остро переживает как потерю экономических благ, которые обеспечивал Интернет, так и возможность общения и поиска информации в режиме онлайн [21, p. 431]. В частности, подруга Ларри Андервуда, Рита Блэйкмур, представитель верхнего слоя среднего класса американского общества, которую Ларри удалось вывести из зараженного и погибающего Нью-Йорка, выражает сожаление не только из-за окружающих их опасностей, но и из-за того факта, что ее айфон новейшей модели превратился в бесполезный агрегат, не способный заполнить ее досуг [21, p. 432].

Важно учесть то обстоятельство, что, согласно как роману, так и сценарию сериала 2020 г., выжившие на Земле люди (составляющие всего один процент от изначального населения планеты), обладающие врожденным иммунитетом, в итоге делятся на два сообщества. Те личности, которые сохранили нравственные ориентиры, концентрируются в г. Боулдер (штат Мэн), месте проживания праведницы Эбигэйл. В свою очередь, индивиды с аморальными поведенческими наклонностями направляются в Лас-Вегас, являющийся штаб-квартирой Рэнделла Флэгга, демонической сущности, инспирировавшей эпидемию. При этом многие из его сторонников относятся к той категории людей, которые не могут преодолеть исчезновение виртуального пространства [23, p. 53]. Наличие большого числа технических специалистов, включающих и экспертов по компьютерному программированию, внушает им

надежду на то, что Флэгг в состоянии вернуть им привычный жизненный уклад, с онлайн-доставками, сетевыми играми и виртуальными коммуникативными площадками [24, p. 219].

В 1980 г. был опубликован роман С. Кинга «Воспламеняющая взглядом», повествующий о супружеской паре Виктории и Эндрю Макги, спасающих свою дочь Чарли (Шарлин) от сотрудников американских спецслужб. Девочка обладает сильными пирокинетическими способностями, т.е. возможностью генерировать исключительно высокую температуру, вследствие экспериментов, проведенных над ее родителями в секретной государственной лаборатории. Снятый в 1984 г. М. Л. Лестером одноименный фильм обрел большую популярность как в США, так и за их пределами.

В 2002 г. Р. Исков снял продолжение — «Воспламеняющая взглядом-2». В данной кинематографической ленте повзрослевшая Чарли Макги и ее партнер Винсент Сфорца выясняют, что опыты, аналогичные экспериментам над родителями девушки, продолжаются в новом исследовательском центре. Макги и Сфорца получают необходимую информацию благодаря взлому центрального компьютерного процессора государственной корпорации, финансирующей данную антигуманную деятельность [11, с. 85].

В данном случае писатель дал позитивную оценку внедрению сцены с компьютерным взломом, прокомментировав свою точку зрения репликой о том, что наличие виртуального пространства и кибернетических технологий существенно упрощает общественный контроль над экспериментальной деятельностью государственных структур и частных компаний, которая зачастую носит характер сомнительный с точки зрения нравственных категорий [19, с. 155].

В 2022 г. режиссер К. Томас создал свою версию «Воспламеняющей взглядом». С. Кинг подверг данную работу острой критике, обвиняя кинематографиста в непонимании его идей, неудачном подборе актеров и излишнем стремлении учитывать т.н. «повесточные тенденции» при формировании образов главной героини и ее матери. Однако писатель дал положительную оценку тому факту, что, согласно сценарию фильма, сотрудникам спецслужб удалось обнаружить скрывающуюся семью Макги благодаря отслеживанию их айфонов. С. Кинг констатировал, что данный факт указывает на фактическую невозможность остаться незамеченным в современном мире, насыщенном высокими технологиями, ко-

торые позволяют вычислить местонахождение человека в любой точке земного шара [13, с. 96].

В 1981 г. С. Кинг завершил работу над романом «Куджо», который посвящен истории одноименной собаки-сенбернара, отличавшейся исключительным дружелюбием, но после укуса летучей мыши подхватившей бешенство и послужившей причиной гибели нескольких человек, включая собственного хозяина.

В одной из ключевых сцен главная героиня романа Донна Трент, которая приехала в находящуюся за городом автомастерскую вместе с четырехлетним сыном Тедди, оказывается заблокированной вместе с ребенком в автомашине, атакованной Куджо, не имея возможности добраться до телефона в доме, чтобы позвонить в полицию. В данном случае проявился прогностический талант писателя: оказавшаяся в катастрофической ситуации Донна сожалеет, что не существует беспроводных телефонов или иных портативных электронных коммуникационных устройств, которые можно было бы использовать при форс-мажорных обстоятельствах [27, р. 202].

В 1983 г. вышел роман автора «Мертвая зона», повествующий о Джоне Смите — молодом американце, получившем в результате тяжелой черепно-мозговой травмы провидческий дар, который, в итоге, позволил ему выявить опасного маньяка Фрэнка Додда и предотвратить приход к власти в США ультраправого политика-демагога Грегори Стилсона, способного развязать ядерный конфликт. Во второй экранизации книги, сериале М. Пиллера и Ш. Пиллера, вышедшем в 2002 г., Грегори Стилсон активно использует интернет-ресурсы для общения со своими избирателями и распространения политической рекламы [7, с. 68]. С. Кинг, в целом сдержанно оценивший работу Пиллеров, позитивно оценил данный факт, отметив, что в ближайшие десятилетия (иными словами — современность) политики радикального толка превратят виртуальное пространство в платформу для распространения своих опасных воззрений, что требует разработки программ, которые были бы способны эффективно противодействовать их деятельности в режиме онлайн, позволяющем одновременно оказывать воздействие на многомиллионную аудиторию [15, с. 84].

В 1986 г. С. Кинг опубликовал роман «Оно», считающийся одной из литературных вершин жанра ужасов. Данное произведение раскрывает такие значимые проблемы, как конфликт поколений,

насилие в семье, сложности процесса социализации в переходном возрасте, способность преодолевать подсознательные страхи, подростковые комплексы, злоупотребление продуктами питания, травля со стороны агрессивных сверстников. В 2015-2017 гг. вышла экранизация романа в двух частях, снятая А. Мускетти. В «Оно-2», снятом в 2017 г., один из главных героев, Майл Хэнлон, находит критически важную информацию об инфернальной сущности, известной под именем Клоун Пеннивайз, на официальном сайте города Дерри, в котором происходят события [3, с. 13].

Кроме того, в данном фильме был использован оригинальный прием, ориентированный на представление роли социальных сетей в современной предпринимательской деятельности. Герой по имени Бен Хэнком в детстве и юности отличался излишним весом, из-за чего подвергался гонениям и испытывал проблемы в отношениях с противоположным полом. В начальных кадрах «Оно-2» зрителям демонстрируют заседание руководства крупной строительной компании, на котором выступает с докладом мужчина, напоминающий Бена в юности — малопривлекательного со стереотипной точки зрения человека. Однако в следующем кадре выясняется, что это его сотрудник, а сам Хэнком, кардинально изменивший внешность в положительную сторону и являющийся собственником данной фирмы, находится в другом городе и ведет совещание посредством формата Zoom [17, с. 75].

В 1992 г. был издан роман «Игра Джеральда», главная героиня которого Джесси Берлингейм, после скоростной смерти супруга Джеральда, оказывается изолированной в дачном домике вдали от населенных пунктов и спасается благодаря силе воли и упорству в достижении цели.

В 2017 г. М. Флэнаган поставил фильм по мотивам данного литературного произведения. В кинокартине Джесси (в исполнении популярной актрисы К. Гуджино), прикрепленная к кровати наручниками, в стремлении спастись долгое время концентрируется на усилиях добраться до лежащего на отдаленном расстоянии айфона, а затем — на старании связаться со службами помощи или знакомыми; только неэффективность этой попытки, связанная с неисправностью мобильного устройства, побуждает Джесси Берлингейм к поиску иных способов спасения, приведших к положительному результату [8, с. 425]. С. Кинг прокомментировал данный сюжетный прием как удачный ход, позволяющий продемонстри-

ровать зрителям большую зависимость от высокотехнологичных устройств, ведущую к беспомощности их владельцев в случае какого-либо сбоя в программе или обычной поломки [26, с. 22].

В этом контексте, что важно отметить, в 2000-х гг., по мере интенсификации развития высоких коммуникационных технологий и онлайн-платформ, данные явления общественной жизни получили непосредственное отражение в творчестве С. Кинга.

При этом мнения критиков и активной части его поклонников разделились. Большая часть из них положительно восприняла данное направление в творчестве любимого автора, другая часть расценила его позицию как технофобскую, критикующую развитие виртуального пространства, в первую очередь — в отношении средств мобильной связи, в особенности наделенных функциями умного устройства, т.е. смартфонов различных моделей [2, с. 60]. Сторонники данной точки зрения, в частности, в качестве основного аргумента указывали на сюжет романа «Мобильник», опубликованного в 2006 г. и описывавшего мир недалекого будущего, в котором произошла техногенная катастрофа. В результате спонтанного концентрированного электронного импульса люди, использующие сотовые телефоны, подвергаются зомбированию и представляют смертельную опасность для окружающих. По понятным причинам, абсолютное большинство представителей человеческого вида утрачивают разум и превращаются в зомби, именуемых в романе мобилоидами. Данная тема получила развитие в экранизации 2016 г., автором которой является Т. Уильямс, а главные роли исполнили выдающиеся голливудские актеры Д. Кьюсак и С. Л. Джексон [29, р. 141].

Отвечая на критику романа на рубеже 2010-2020-х гг., С. Кинг отметил, что не является противником сотовой связи и смартфонов как таковых. Писатель подчеркнул, что в «Мобильнике» стремился указать на максимальную вовлеченность людей в виртуальное общение и во время свободного времяпровождения, и в период учебного процесса, и при исполнении трудовых функций, при постепенном игнорировании реального мира или его восприятию в качестве второстепенной и обременительной среды обитания [22, р. 196]. Однако его оппоненты в качестве дополнительного довода концентрируют внимание на повести 2020 г. «Телефон мистера Харригэна», включенной в сборник «Будет кровь». В соответствии с сюжетной линией, главный герой, подросток по имени

Крэйг, на похоронах своего пожилого друга и работодателя Джона Харригэна кладет ему в карман подаренный ранее айфон, с которым Харригэна опускают в могилу. Позднее, под воздействием эмоций, Крэйг отправляет голосовые сообщения на данное мобильное устройство с целью поделиться своей проблемой в лице преследующего его хулигана. В результате противник главного героя совершает суицид. Через несколько лет после аналогичного звонка на телефон, погребенный вместе с владельцем в могиле, совершает самоубийство автомобилист, явившийся причиной гибели любимого педагога подростка. После этого Крэйг удаляет номер Харригэна из своих контактов [20, р. 39].

Лица, обвиняющие С. Кинга в технофобии, полагают, что основной смысл повести, как и экранизации Д. Л. Хэнкока (с Д. Сазерлендом и Д. Мартеллом в главных ролях), заключается в утверждении, что использование смартфонов оказывает вредоносное воздействие на их обладателей [25, р. 88]. В ответ на это обвинение писатель констатировал, что такая претензия ошибочна. Ключевая идея произведения заключается в том, что неопределенно-длительное использование смартфонов способно привести к искаженному восприятию реальности, агрессивным намерениям и совершению алогичных поступков. Те же, кто применяет данное умное устройство в разумных пределах, сохраняют разум и не представляют угрозы для социума [22, р. 273].

Кроме того, в романе 2005 г. «Парень из Колорадо» девяностолетний журналист Винсент Тиг в беседе со своей молодой коллегой Стефани Маккэн выражает удовлетворение тем, что использует в своей профессиональной деятельности новейшую модель стационарного компьютера «Макинтош» и сотовый телефон последней модели. Тиг подчеркивает, что наличие данных устройств в прошедших десятилетиях в значительной степени повысило бы продуктивность его работы [25, р. 268].

В 2014-2016 гг. С. Кинг создал «Трилогию Билла Ходжеса» об отставном полицейском-пенсионере, работающем в качестве частного детектива. Романы «Мистер Мерседес», «Кто нашел, берет себе» и «Пост сдал» в основном ориентированы на отображение воззрений писателя относительно вопросов сетевого троллинга, компьютерофобии и интернетофобии, хакерства, сетевых вирусов и специфики поведенческой модели компьютерных программистов и иных работников сферы высоких технологий. В столь же

большой степени изучаемая проблема рассмотрена в экранизации — сериале Д. Э. Келли 2017-2019 гг. [8, с. 426].

Однако, как представляется авторам данной статьи, освещение рассматриваемой темы в указанной выше трилогии, с учетом ее многоаспектности и смысловой глубины, является объектом отдельного научного исследования.

Заключение

На основе изучения рассмотренных в данной статье материалов следует констатировать, что фантастическая вселенная С. Кинга на протяжении нескольких десятилетий является площадкой изучения эволюции виртуального пространства и инновационных коммуникативных устройств.

Данный автор, несмотря на необоснованные обвинения (равно как и в отношении кинематографистов, работающих над экранными адаптациями его романов, рассказов и повестей) в отсутствии смысловой нагрузки в своем творчестве, с самого начала своей феноменальной литературной карьеры проявляет интерес к развитию системы электронных технологий и ее восприятию в общественном сознании; отображение изучаемой проблемы в его творчестве и в деятельности сотрудничающих с ним деятелей кинематографического искусства в начале 1970-х — начале 2020-х гг. характеризует и эволюцию восприятия виртуального пространства и инновационных вычислительных и коммуникационных устройств в общественном мнении.

В произведениях С. Кинга и экранизирующих его книги кинорежиссеров раскрываются причины негативного отношения к высоким технологиям со стороны религиозных фанатиков во второй половине XX столетия, указывается на необходимость общественного контроля за социальными сетями, перед человеческим сообществом ставится вопрос об этичности и степени безопасности неограниченного доверия искусственному интеллекту в военной, экономической и социальной сферах жизни мирового социума.

В то же время следует принять во внимание и то важное обстоятельство, что обвинения в технофобии и враждебности по отношению к высокотехнологичным сотовым телефонам и социальным сетям, предъявляемые создателям фантастической вселенной С. Кинга, являются необоснованными и не соответствующими ре-

альности. В данной многовекторной франшизе лишь высказывается предупреждение лицам, чрезмерно злоупотребляющим данными технологиями.

Литература

1. Афанасьев А. Н. Дети Интернета. Что они смотрят и кто ими управляет? М.: Наше Завтра, 2021. 314 с. (Afanasiev A. N. Children of the Internet. What are they watching and who controls them? M.: Nashe Zavtra, 2021. 314 c.)
2. Баева Л. В. Образ киберчеловека в современной науке и культуре // Философские проблемы информационных технологий и киберпространства. 2015. № 1(9). С. 56-69. (Baeva L. V. The image of a cyberman in modern science and culture // Filosofskie problem informatsionnykh tekhnologiy i kiberprostarnstva. 2015. № 1(9). S. 56-69.)
3. Баранов Б. Н. Социальные сети // Управление. 2020. № 2. С. 12-14. (Baranov B. N. Social networks // Upravlenie. 2020. № 2. S. 12-14.)
4. Берн Э. Игры, в которые играют люди. Люди, которые играют в игры. М.: Эксмо, 2023. 288 с. (Bern E. The games that people play. People who play games. M.: Eksmo, 2023. 288 s.)
5. Варшавская Ю. Л. Космические киноужасы // Литературная газета. 1977. № 36. С. 15. (Varshavskaya Yu. L. Space movie scenes // Literaturnaja gazeta. 1977. № 36. S. 15.)
6. Горохов П. А. Фантастика Стивена Кинга: традиции и новаторство // Вестник ОГУ. 2004. № 11. С. 179-184. (Gorokhov P. A. Stephen King's fiction: traditions and innovation // Vestnik OGU. 2004. № 11. S. 179-184.)
7. Гошило Е., Гошило Б. Выцветание красного: бывший враг времен холодной войны в русском и американском кино 1990-2005 годов. СПб.: Academic Studies Press, 2020. 480 с. (Goshilo H., Goshilo B. Fade from Red. The Cold War Ex-Enemy in Russian and American Film 1990-2005. SPb: Academic Studies Press, 2020. 480 c.)
8. Давыдова М. М. Концепт «самоубийство» и средства его репрезентации в романе С. Кинга «Пост сдал» // Российский гуманитарный журнал. 2021. Т. 10. № 6. С. 424-428. (Davydova M. M. The concept of «suicide» and the means of its representation in S. King's novel «The Post has passed» // Rossiyskiy gumanitarnyy zhurnal. 2021. T. 10. № 6. S. 424-428.)
9. Дустмурадов Т. А. Индивидуально-авторский концепт «страх» в произведениях Эдгара По и Стивена Кинга // Аллея науки. 2018. Т. 2. № 5(21). С. 563-567. (Dostmuradov T. A. The individual author's concept of «fear» in the works of Edgar Poe and Stephen King // Alleya nauki. 2018. T. 2. № 5(21). S. 563-567.)
10. Козьмина Е. Ю. Фантастика начала XX века в жанровом освещении // Научный диалог. 2017. № 6. С. 148-157. (Kozmina E. Y. Fiction of the early twentieth century in genre lighting // Nauchnyj dialog. 2017. № 6. S. 148-157.)

11. Кошкин А. П., Мельков С. А. Герой в России: кто он? (анализируя современный кинематограф) // Национальные и образовательные проблемы гражданской защиты. 2010. № 1. С. 83-90. (Koshkin A. P., Melkov S. A. Hero in Russia: who is he? (analyzing modern cinema) // Nacional'nye i obrazovatel'nye problemy grazhdanskoj zashhity. 2010. № 1. S. 83-90.)
12. Манахов А. А. Фантастика и фэнтези: стилистические различия // Science and education. 2023. Т. 2. № 4. С. 1279-1283. (Manakhov A. A. Science fiction and fantasy: stylistic differences // Science and education. 2023. T. 2. № 4. S. 1279-1283.)
13. Окавка Л. Б. Социальный статус кино как феномена художественной культуры // Гуманитарные исследования в Восточной Сибири и на Дальнем Востоке. 2021. № 1. С. 94-100. (Okavka L. B. The social status of cinema as a phenomenon of artistic culture // Gumanitarnye issledovaniya v Vostochnoj Sibiri i na Dal'nem Vostoke. 2021. № 1. S. 94-100.)
14. Попов Д. А. Творчество Стивена Кинга как художественное воплощение психоаналитических идей // Мир науки, культуры, образования. 2014. № 2(45). С. 281-283. (Popov D. A. Stephen King's work as an artistic embodiment of psychoanalytic ideas // Mir nauki, kul'tury, obrazovaniya. 2014. № 2(45). S. 281-283.)
15. Разлогов К. Э. Утопия и антиутопия в мировом кинематографе // Международный журнал исследований культуры. 2012. № 4 (9). С. 112-115. (Razlogov K. E. Utopia and dystopia in world cinema // Mezhdunarodnyj zhurnal issledovanij kul'tury. 2012. № 4 (9). S. 112-115.)
16. Роугек Л. Сердце, в котором живет страх. Стивен Кинг: жизнь и творчество. М.: Астрель, 2011. 416 с. (Rogek L. The heart in which fear lives. Stephen King: Life and work. M.: Astrel', 2011. 416 s.)
17. Филистова Н. Ю., Сакова А. С. Вербальная реализация концептов «fear» и «madness» в литературе жанра ужасов (на материале романа Стивена Кинга «The Shining») // Актуальные вопросы современной филологии и журналистики. 2019. № 1. С. 73-80. (Filistova N. Yu., Sakova A. S. Verbal realization of the concepts of «fear» and «madness» in the literature of the horror genre (based on the novel by Stephen King «The Shining») // Aktual'nye voprosy sovremennoy filologii i zhurnalistiki. 2019. № 1. S. 73-80.)
18. Хрюкин Д. А. «Сотворение мира»: способы образного моделирования в научно-фантастическом кино // Артикульт. 2014. № 1(13). С. 81-87. (Khryukin D. A. "Creation of the world": methods of figurative modeling in science fiction cinema // Artikul't. 2014. № 1(13). S. 81-87.)
19. Шалимова И. В., Годунова К. Е. Концепты «Imprisonment» (заключение) и «Madness» (безумие) в творчестве Стивена Кинга // Вестник КемГ У. 2013. Т. 2. № 4 (56). С. 153-157. (Shalimova I. V., Godunova K. E. The concepts of «Imprisonment» (conclusion) and «Madness» (madness) in the works of Stephen King // Vestnik KemG U.

2013. Т. 2. № 4 (56). S. 153-157.)
20. Clarke W. H. It ain't the guns Mr. King: a rebuttal to Stephen King's essay about guns. Augusta: Lime Line Books, 2013. 107 p.
 21. Hafdahl M., Florence K. The Science of Stephen King: The Truth Behind Pennywise, Jack Torrance, Carrie, Cujo, and More Iconic Characters from the Master of Horror. New-York: Skyhorse, 2020. 544 p.
 22. King S. On Writing: A Memoir of the Craft. Scribner: Scribner edition, 2020. 320 p.
 23. Markham A. Stephen King books in order. Scribner: Scribner edition, 2024. 137 p.
 24. Olson D. Children and Childhood in the Works of Stephen King (Children and Youth in Popular Culture). Lanham: Lexington Books, 2020. 353 p.
 25. Roch M. Fun is Fun and Done is Done: Stephen King Connections and Easter Eggs. Ventura: MicRo Readers, 2022. 331 p.
 26. Soliyev F. Semantic and discursive analysis of detective fiction «Mr. Mercedes» by Stephen King // Science and innovation. 2024. № 3(9). P. 22-25.
 27. Spignesi S. Stephen King, American Master: A Creepy Corpus of Facts About Stephen King & His Work. New-York: Permuted Press, 2018. 432 p.
 28. Spignesi S. The Essential Stephen King: A Ranking of the Greatest Novels, Short Stories, Movies, and Other Creations of the World's Most Popular Writer. Sacramento: New Page Books, 2001. 359 p.
 29. Wiater S. The Complete Stephen King Universe: A Guide to the Worlds of Stephen King. New-York: St. Martin's Griffin, 2006. 264 p.

УДК 004.8 + 004.93

DOI 10.17726/philIT.2024.2.6



Почему трансформеры не мыслят, как люди?

Александр Б. Хомяков,

магистр физических наук

Санкт-Петербург, Россия

alexander.xom@gmail.com

Аннотация. Большие языковые модели в виде чат-ботов очень правдоподобно имитируют диалог как всезнающий собеседник и поэтому получили широкое распространение. Но даже в чат-боте Google Gemini не советуют доверять тому, что напишет чат-бот, и просят проверять его ответы. В данном обзоре будут проанализированы различные типы ошибок LLM, такие как проклятие инверсии, обработка чисел и др., чтобы выявить их причины. Такой анализ привел к выводу об общих причинах ошибок, заключающихся в том, что трансформеры не обладают глубокой аналогией, абстракцией и избирательностью контента, учитываемого в вычислении ответа (inference). Но наиболее важным выводом является то, что трансформеры, как и другие нейросети, построены по концепции обработки входного сигнала, что создает сильную зависимость от нерелевантной информации, которую не может компенсировать слой внимания трансформера. Концепция нейросетей была заложена в 1950-х идеей перцептрона Ф. Розенблата и не учитывала тех достижений когнитивной психологии, которые появились позже. Согласно же конструктивистской парадигме, входной слой (или перцепция) является только способом проверки правильности сконструированной предиктивной модели для возможных ситуаций. Это же служит причиной самой большой проблемы трансформеров, называемой галлюцинациями. И устранение ее возможно только при изменении архитектуры нейросети, а не за счет большего количества данных в обучении.

Ключевые слова: LLM; трансформеры; мышление; аналогия; когнитивная психология; перцептивный цикл; галлюцинации.

Why don't transformers think like humans?

Alexander B. Khomyakov,

Master of Physical Sciences

Saint-Petersburg, Russia

alexander.xom@gmail.com

Abstract. Large language models in the form of chatbots very realistically imitate a dialogue as an omniscient interlocutor and therefore have become widespread. But even Google in its Gemini chatbot does not recommend trusting what the chatbot will write and asks to check its answers. In this review, various types of LLM errors such as the curse of inversion, number processing, etc. will be analyzed to identify their causes. Such an analysis led to the conclusion about the common causes of all errors, which is that transformers do not have deep analogy, hierarchy of schemes and selectivity of content taken into account in the inference. But the most important conclusion is that transformers, like other neural networks, are built on the concept of processing the input signal, which creates a strong dependence on superficial noise and irrelevant information that the transformer's attention layer cannot compensate for. The concept of neural networks was laid down in the 1950s by the idea of F. Rosenblatt's perceptron and did not take into account the achievements of cognitive psychology that appeared later. According to the constructivist paradigm, the input word (or perception) is only a way to check the correctness of the constructed predictive model for possible situations. This is the cause of the biggest problem of transformers, called hallucinations. And its elimination is possible only by changing the architecture of the neural network, but not by increasing the amount of data in training.

Keywords: LLM; transformers; thinking; analogy; cognitive psychology; perceptual cycle.

Введение

Современные трансформеры требуют огромных вычислительных ресурсов. Архитектура трансформера масштабируется квадратично по мере увеличения длины входной последовательности. То есть, когда длина последовательности, обрабатываемой трансформером (количество слов в промте), увеличивается, требуемые для обработки вычисления увеличиваются на эту величину в квадрате и быстро становятся неподъемно огромными. Это приводит, например, к невозможности сильно увеличивать окно промта, а его размер в действительности увеличивается за счет алгоритмов сжатия векторов, что неизбежно вызывает потерю информации из промта, особенно фактов, следствием чего является неточность ответа и галлюцинации. Таким образом, трансформеры в действительности не могут работать с длинными последовательностями.

Кроме того, на прошедшей в ноябре 2024 г. конференции NeurIPS Илья Суцкевер заявил, что в мире нет больше данных, чтобы обучить трансформеры в надежде на эффект масштаба (эффект улучшения способностей трансформера при обучении на больших данных). «У нас только один интернет», — сказал он и подчеркнул, что трансформеры, чтобы решить проблему данных, не должны обучаться так, как сейчас [1]. Это ограничение не только данных, но и архитектуры. Это известный факт того, что трансформеры не могут обучаться во время использования, как это делает человек. Но он не ответил на вопрос, а как же должна измениться архитектура нейросети, указав только на то, что они должны стать агентами в среде, как и люди. Необходимость невообразимого количества данных и невозможность обучаться во время использования говорит о том, что трансформеры в своей архитектуре столкнутся с преградой в дальнейшем развитии.

Цель исследования

Но это далеко не все проблемы, которые не преодолимы трансформерами без помощи людей. К ним относят и проблемы с логикой и с вычислениями в больших последовательностях, и следование правилам, и, конечно же, галлюцинации. Такие примеры и будут проанализированы в статье. Отсюда следует все чаще звучащий тезис о том, что необходимо искать новую архитектуру для создания поистине интеллектуальных систем. Данная статья посвящена поиску тех архитектурных принципов, которые могут приоткрыть путь к новой архитектуре. И в этом нам помогут как раз те ошибки трансформеров, которые описаны исследователями. Они прямо указывают на принципиальные недостатки, причины которых могут быть как раз в отсутствии необходимых функций и структур нейросети. И цель исследования состоит в анализе таких ошибок «мышления» трансформеров, чтобы предположить, каких способностей не хватает трансформерам.

Надо отметить, что большинство тестовых заданий, которые приведены ниже, в итоге попадают в Интернет и таким образом оказываются в обучающих данных новых языковых моделей, что позволяет им успешнее их решать. Это называется «утечкой тестовых данных» (data leakage). Более того, компании, создающие большие языковые модели, нанимают десятки тысяч ассессоров, которые находят ошибки трансформеров и дообучают их (обу-

чение за счет человеческой обратной связи, RLHF). В результате нейросеть может показывать высокие результаты на тестах не благодаря своим способностям или пониманию заданий, а из-за того, что она натренирована их решать правильно. Такой подход к решению проблем LLM еще раз подчеркивает, что они являются очень хорошими имитаторами с аппроксимацией, но не способны сами прийти к решению.

Материалы и методы исследования

Трансформеры называют «черными ящиками», так как интерпретация того, как и почему они генерируют тот или иной ответ, оставалась загадкой. Но с недавнего времени появились инструменты интерпретации LLM, которые позволяют проследить, какие группы нейронов имели большее влияние на результат на выходе трансформера. И удалось также идентифицировать эти группы по тематикам, за которые они отвечают. Такие инструменты появились у Anthropic, у OpenAI. Совсем недавно появился стартап Transluce (<https://transluce.org>), который создал инструмент Monitor (<https://monitor.transluce.org>), позволяющий посмотреть, какие группы нейронов активируются при вычислении ответа на введенный промт в Llama-3.1-8b-instruct. Это дает возможность увидеть, что повлияло на ответ, какие тематики внесли какой вклад в ответ и на какие слова в ответе больше всего активированы. При этом можно «отключить» или, наоборот, «активировать» эти группы, чтобы посмотреть, как изменится ответ. Это совершенно новый инструмент исследования, который делает трансформеры не такими черными ящиками, как считалось ранее, и позволяет изучать их работу, их «интеллект».

Что больше, 9.8 или 9.11

Например, известная проблема для LLM, заключающаяся в сравнении чисел 9.8 и 9.11. Модели LLM традиционно считают, что 9.11 больше. Так в чем же дело? Мы посмотрели в Монитор Transluce. На рис. 1 видно, что модель (в данном случае Llama-3.1), когда пытается сравнить числа 9.8 и 9.11, активирует несколько неожиданных кластеров: например, про атаку 11 сентября (9.11) и гравитационную константу (9.8), а также нумерацию стихов из Библии, где действительно 9.11 идет раньше 9.8. Вероятно, из-за того, что эти темы появлялись в обучающих данных Llama очень

часто, числа 9.8 и 9.11 перестают восприниматься ею как обычные числа: она воспринимает их как другой вид объектов (даты, константы, параграфы).

Model Chat (Llama-3.1 8B Instruct)

Try my own prompt!

I found some neurons that fired highest on the input. I grouped their behaviors into these clusters:

- 9/11 terrorist attacks, geopolitical and economic events
6 neurons matching
- cooking ingredients, recipes, food guides
5 neurons matching
- statistical measurements, medical dosages
5 neurons matching

High-Activation Neurons
Showing neurons that fire highly on your prompt.

Quickly search neuron descriptions by keyword...

ID	Act / Top Nile	Explanation
L27/ N8074	4.2870	the term <code><end_header_id></code> , indicating a comparison or relationship
L11/ N1924	3.5143	links that contain specific domain identifiers <code><math></code> , <code><code></code> , <code><math></code> , and <code><code></code>
L27/ N1772	3.3722	activation occurs on specific numeric tokens; numbers at the beginning of a context, particularly in mathematical or programming contexts, particularly ones that repeat or pertain to a calculation or structure.
L30/ N3886	3.1905	expressions of comparison including "is" or "not" with numbers or concepts of numerical value

Page 1 of 104

Но при отключении этих групп нейронов с явно нерелевантными вопросу темами удалось получить верный ответ. Отсюда можно сделать вывод, что трансформеры при определении ответа не могут выбрать ключевую тему и далее не учитывать другие при поиске ответа. Архитектурно трансформер так устроен, что вычисление производится по всем имеющимся параметрам нейросети без избирательности.

Алиса в стране чудес

В статье «Alice in Wonderland: Simple Tasks Showing Complete Reasoning Breakdown in State-Of-the-Art Large Language Models» [4] авторы задают вопрос трансформерам: «У Алисы есть Н братья и у нее также есть М сестры. Сколько сестер у брата Алисы?». Они сделали тестовый набор аналогичных задач с разной длиной контекста и провели испытания лучших больших языковых моделей на начало июня 2024 года. Ни одна модель не смогла справиться с заданиями полностью. Только GPT-4 в некоторых случаях при пошаговом объяснении (step-by-step prompt) дает правильный ответ, но ошибается при более сложной формулировке задачи. А люди (с определенного возраста) легко справляются с ней.

В статье Ben Hagag [7] дается анализ того, как LLM Lamma 3.1 отвечает на этот вопрос с помощью монитора Translucide. Автор

подчеркивает, что Llama изначально не дает правильный ответ, как и большинство других моделей. Но при отключении активированных нейронов явно неуместных тем (химических соединений и реакций) и усилении активации нейронов темы «гендерные отношения» удалось получить правильный ответ.

Мы взяли другой вопрос для тестирования, на котором ошибаются и такие модели, как ChatGPT-4o. «У Пети братьев и сестер поровну. Кого в семье больше, сестер или братьев». Llama отвечает, как и ChatGPT, что число сестер и братьев в семье одинаково, что ошибочно. Хотя наибольшие активации получают правильные нейроны, отвечающие на семейные отношения. И вторым по количеству нейронов активируется тема математических отношений, что тоже валидно вопросу, так как речь о сравнении количества. Но это не помогает сделать правильный ответ. И попытки отключить или сильнее активировать какие-либо группы нейронов не дали лучшего результата.

Model Chat (Llama-3.1 8B Instruct)

Try by own prompt!

I found some neurons that fired highest on the input. I grouped their behaviors into these clusters:

- familial relationships, sibling dynamics, parenting, 71 neurons matching
- comprehensive mathematical concepts, operations, education 62 neurons matching
- diverse scientific and technical terms 16 neurons matching

Activation Mode **Attribution Mode** **High-Activation Neurons**

Showing neurons that fire highly on your prompt.

Quickly search neuron descriptions by keyword...

Neuron ID	Score	Description
L14/ N693	3.2727	the word "that" following a comparative context, along with "having," "become," or similar structures emphasizing comparison or equality.
L13/ N1130	3.0863	conjunctions (e.g., "and") appearing before clarifications or additional details
L31/ N12082	3.0270	the words "many," and "too" when discussing siblings or family relationships
L15/ N2482	2.9778	phrases indicating equality (e.g., "is," "are") especially in mathematical or definitional contexts
L22/ N808	2.9159	"in" when indicating quantity or count references.
L13/		mentions of joint training programs or initiatives (e.g., "jointly commissioned," "joint service") and cooperation

Previous Page 1 of 208 Next

Так в чем проблема решения такой задачи как когнитивного процесса? Представляется, что проблема не только в активации лишних тематических групп нейронов, так как две самые активные группы кажутся достаточными для правильного ответа — семейных отношений и операций сравнения, но и в чем-то ещё, чтобы ответ был верным. Человек выполняет эту задачу в несколько шагов, создавая сначала пространственную картину отношений: вот стоит Петя и вот рядом его, например, три брата и три сестры. Но он тогда четвертый брат в семье, что очевидно при таком пред-

ставлении. И похоже, что таких промежуточных представлений у трансформеров нет. Они пытаются в один шаг найти эвристику, наиболее вероятную для такого запроса. И она неверная.

Проклятие инверсии

Это пример отсутствия у LLM возможности сделать из послылки $A \rightarrow B$ вывод об отношении $B \rightarrow A$, которая в оригинальной статье Lukas Berglund and etc. [6] демонстрируется вопросами “Who is Tom Cruise’s mother? [A: Mary Lee Pfeiffer]” and the reverse “Who is Mary Lee Pfeiffer’s son?”. На второй вопрос LLM не дает ответа, хотя он очевиден. При рассмотрении активации тематических групп в мониторе (рис. 3) можно заметить, что активируется более всего тема ролей в семье и отношений, что верно. Но почему же ответ все равно неверный?

The figure displays two screenshots of the Translucence Model Investigator interface, showing the analysis of neuron activation for the Llama-3.1-8B-Instruct model.

Top Screenshot: The interface shows the prompt "Who is Tom Cruise's mother?" and the model's response "Tom Cruise's mother is Mary Lee Pfeiffer. She was a special education teacher and a homemaker. Mary Lee Pfeiffer was born on May 2, 1936, and passed away on June 16, 2022, at the age of 86." The "High-Activation Neurons" section shows clusters of neurons that fired highest on the input, grouped by their behaviors into three clusters: specific dates, medical terminology, and film industry. The "Activation Mode" is set to "High-Activation Neurons".

Bottom Screenshot: The interface shows the prompt "Who is Mary Lee Pfeiffer's son?" and the model's response "I don't have information on who Mary Lee Pfeiffer's son is. Is there anything else I can help you with?". The "High-Activation Neurons" section shows clusters of neurons that fired highest on the input, grouped by their behaviors into three clusters: familial roles, scientific terminology, and military operations. The "Activation Mode" is set to "High-Activation Neurons".

В нейropsychологии известны нарушения в понимании родственных связей, возникающие при повреждении третичных ассоциативных зон в верхней части задневисочной зоны на пересечении с затылочной и теменной. Эта зона отвечает за абстрактные понятия и отношения, включая математические [5]. А. Лурия отмечает, что «общим для всех этих конструкций является то, что ни одну из них нельзя выразить в наглядном образе; все они, таким образом, различными способами кодируют не наглядные, а логические отношения» [5]. Это наталкивает на гипотезу о том, что трансформерам не хватает схем абстрактных обратимых отношений (инверсии), которые особенно характерны для математических операций, пространственных и родственных отношений. Например, если $3+2=5$, то $5-2=3$. Мы выучиваем не конкретные примеры, а абстрактную схему прибавления и вычитания как обратную. Именно такую абстракцию не может извлечь трансформер, несмотря на обильное обучение на примерах.

Зависимость от формулировок

Мы также решили проверить в мониторе известную проблему трансформеров на зависимость ответа от формулировок запроса в промте. В результате ответ может быть совершенно другой при перефразировке того же вопроса. Мы предложили две формулировки силлогизма: более известную и неизвестную трансформеру (см. рис. 4 и 4а). В первом случае ответ был верен, так как речь шла о Сократе. Трансформер правильно определил философскую тематику — это наиболее активная группа нейронов. Далее активировались нерелевантные темы математических отношений и религиозных текстов, но это не помешало дать правильный ответ. В случае же неизвестной формулировки того же силлогизма трансформер дал неверный ответ. Упоминание бога сделало доминирующей тему «теологического дискурса», а также мифологии и фантастики, что не позволило трансформеру дать правильный ответ. Трансформер при этом не увидел в тексте силлогизм по аналогии, что является ключевой ошибкой. Введение в текст дополнительных слов, не относящихся к вопросу, также нередко меняет ответ трансформеров на неверный. Возникает вопрос, насколько в действительности трансформеры хорошо понимают задание, если их понимание легко сломать переменой слов.

Model Chat (Llama-3.1 8B Instruct)

`<[start_header_id]>[user]<[end_header_id]> <[start_header_id]>[assistant]<[end_header_id]>`

Socrates is a man, men are mortal, is Socrates mortal?

This is a classic example of a syllogism, a form of logical argument that uses deductive reasoning to arrive at a conclusion. The argument is as follows:

- Socrates is a man.
- All men are mortal.
- Therefore, Socrates is mortal.

This argument is valid because it

Continue chatting with the model...

Llama-3.1-8B-Instruct Clear chat Regenerate Send

Steering

Add or remove concepts from the model's computation.

Type a concept to steer with...

Strengthen Suppress

Translucence Model Investigator

Try my own prompt!

I found some neurons that fired highest on the input. I grouped their behaviors into these clusters:

philosophical concepts, ancient philosophers, debates.. 79 neurons matching	mathematical proofs, concepts, and logic 73 neurons matching	religious texts, figures, and concepts 38 neurons matching
---	--	--

Activation Mode Attribution Mode High-Activation Neurons

Showing neurons that fire highly on your prompt.

Quickly search neuron descriptions by keyword...

L3/N8154	4.9436	activations on the name "Quasquet" appearing in philosophical contexts as well as variations in spelling of "Quivote" and "Socrat".
L4/N3641	4.5067	presence of a specific symbol indicating continuation or additional content
L9/N401	4.3567	logical correctness and analysis of categorical statements and their implications
L2/N7545	4.1975	the word "Plato" or its variations (e.g., Plato, Platonic, platonism, Platon) activating in various contexts, often related to philosophy, morality, or psychology
L21/N5430	3.5976	sentences containing "think" and its translations (e.g., "ypomnu") and punctuation (e.g., ",")
L11/N1034	3.5714	links that contain specific domain delimiters .pub, .news, .art, and .cs.

Previous Page 1 of 246 Next

Model Chat (Llama-3.1 8B Instruct)

`<[start_header_id]>[system]<[end_header_id]> <[start_header_id]>[user]<[end_header_id]> <[start_header_id]>[assistant]<[end_header_id]>`

Vayya is a god, all are gods, is Vayya a human being?

No, Vayya is not a human being. The statement "Vayya is a god, all are gods" implies that Vayya is a deity, not a human. The phrase "all are gods" suggests that everyone, including Vayya, is considered a god, which means they are not

Continue chatting with the model...

Llama-3.1-8B-Instruct Clear chat Regenerate Send

Steering

Add or remove concepts from the model's computation.

Type a concept to steer with...

Strengthen Suppress

Translucence Model Investigator

Try my own prompt!

I found some neurons that fired highest on the input. I grouped their behaviors into these clusters:

divine, spiritual, theological discussions 86 neurons matching	mythology, gods, cultural narratives 46 neurons matching	fantasy and supernatural elements, diverse settings 25 neurons matching
--	--	---

Activation Mode Attribution Mode High-Activation Neurons

Showing neurons that fire highly on your prompt.

Quickly search neuron descriptions by keyword...

N1962		
L19/N1811	2.3311	tokens indicating reference or direction to previous material (e.g., "supra," "and," "see," "the subsequent," "in")
L14/N8	2.3017	the token "as" indicating direct speech or citation from users
L23/N3155	2.2979	use of the token "as" in contexts involving divine or angelic themes; reference to characters with magical or divine traits; mentions of auras or divine light
L1/N5106	2.2945	variants of "deity" (degg) and "deities" (degag) in various contexts
L12/N12913	2.2788	conversational markers, logical contradictions, undefined states, and contrasting qualities

Previous Page 1 of 203 Next

Эти факты означают, что поверхностный слой (вход) нейросети является определяющим для всей последующей цепочки распространения активации и для получения ответа. Понимание человека более устойчиво к перефразировкам. Человек прежде всего воспринимает смысл вопроса, благодаря чему мы легко определяем одинаковый по смыслу вопрос, даже если он выражен иначе. Но что это значит, быть одинаковым по смыслу? Предполагается, что это возможность установить аналогию между высказываниями благодаря функции аналогии, описанной в нашей предыдущей статье [8]. Это означает, что мы мыслим не конкретными выраженными в тексте последовательностями, а некими концептами как совокупностями аналогичных выражений, но которые не привязаны, не зависят от конкретного выражения в поверхностной структуре (тексте). Мы как бы оторваны от него и мыслим на концептуальном уровне при помощи аналогий.

Концепция нейросетей основана на классическом перцептроне. Он был описан Ф. Розенблаттом в 1957 году [9]. В то время в науке царила кибернетика, зарождались компьютерные науки, преобладающим представлением в них была передача и обработка информации, основы его были заложены К. Шенноном в 1948 году. Это подразумевает, что информация содержится вовне нас, мы воспринимаем при помощи перцепции (входной слой нейросетей) и обрабатываем (внутренние слои) для получения результата (выходной слой). На этом основаны все нейросети, но это уже неверно с точки зрения современной когнитивной науки. Мы не воспринимаем концепты извне, а генерируем их в себе. Перцепция же является всего лишь способом выбора и подтверждения той или иной концепции. Так утверждается в многочисленных гипотезах, таких как гипотеза контролируемых галлюцинаций [10] К. Фристана, гипотеза мультимодального пользовательского интерфейса (MUI) Д. Хоффмана [11] и другие конструктивистские концепции в психологии и философии науки. Трансформеры по своей архитектуре не соответствуют этим современным концепциям.

Какой же может быть перспективная архитектура интеллектуальных систем? В этом поиске стоит обратить внимание на перцептивный цикл У. Найсера [2]. Он отличается тем, что совмещает в себе перцептивный подход с конструктивным. В основе перцептивного цикла стоит схема как некий функционал для возможности воспринимать информацию извне. Схема служит для организации информации со всеми ее возможными вариациями, что близко к тому, что утверждается нами про концепты по аналогии выше. Перцепция служит для выбора варианта схемы, но, если ее сочетание не укладывается в какую-либо схему, происходит, по Найсеру, модификация схем. В этом и заключается работа интеллекта, что схоже также с процессами ассимиляции и аккомодации Пиаже. Трансформеры просто предсказывают следующее слово, даже не пытаясь сверять свои концепции с тем, что написано в промте и выдано ими. Но выстраивание такой архитектуры является делом будущего.

Заключение

Приведенные выше примеры неверной работы трансформеров при ответах — это далеко не все примеры, где они ошибаются. Известна проблема понимания отрицания, так как оно реже встре-

чается в обучающих текстах. Это может быть связано с тем же, что указано в разделе про проклятие инверсии: трансформер не может освоить абстрактную обратимую схему утверждения-отрицания. Также известна проблема большой уверенности трансформеров, которые не задают уточняющих вопросов, не сверяются с другими знаниями, имеющимися у них (что выясняется при наводящих вопросах). Это часто является причиной галлюцинаций в ответах трансформеров. У трансформеров возникают проблемы с умножением больших чисел и их сортировкой. Если попросить перемножить большие числа, особенно если числа написаны текстом, даже при рассуждении по шагам трансформеры часто делают ошибки.

Поэтому наша статья не претендует на всеобъемлющий обзор. Она только начало исследований, показывающее возможности, которые открываются для этого такими инструментами, как монитор Translucence. Но уже данное исследование позволяет выдвинуть некоторые гипотезы относительно того, чего же не хватает трансформерам, чтобы приблизиться к интеллекту человеческого уровня не за счет огромного числа примеров, выученных трансформерами на этапе тренировки.

Во-первых, это отсутствие избирательности выбираемых для ответа тем, соответствующих теме вопроса. Это приводит к влиянию нерелевантных групп нейронов и ошибочным ответам.

Во-вторых, трансформерам не хватает общих и доминирующих над контекстом абстракций с обратимыми отношениями, как то: математические, пространственные и родственные отношения, а также отрицания. Это приводит к проблемам с логическими выводами из известных трансформерам фактов.

В-третьих, построение в парадигме обработки информации и полная зависимость от входа (поверхностной структуры) приводят к возможности неверного ответа при иной постановке вопроса.

Общий вывод: нынешние нейросети, включая трансформеры, не являются еще интеллектом уровня человека, и на сегодня очевидно, что для его достижения необходим поиск новой архитектуры интеллектуальных систем.

Литература

1. Robison K. OpenAI cofounder Ilya Sutskever says the way AI is built is about to change // The Verge, Dec 14, 2024. <https://www.theverge.com/2024/12/13/24320811/what-ilya-sutskever-sees-openai-model-data-training>.

2. Найссер У. Познание и реальность. М.: Прогресс, 1981. С. 42-43. 230 с. (Neisser W. Cognition and Reality. M.: Progress, 1981. P. 42-43. 230 p.)
3. Mitchell M. How do we know how smart ai systems are? // Science, 381(6654). <https://www.science.org/doi/10.1126/science.adj5957>.
4. Nezhrina M. Alice in Wonderland: Simple Tasks Showing Complete Reasoning Breakdown in State-Of-the-Art Large Language Models. <https://arxiv.org/html/2406.02061v1#bib.bib12>, 04 Jun 2024.
5. Лурия А. Р. Основы нейропсихологии: учеб. пособие. М.: Издательский центр «Академия», 2003. 384 с., с. 123-126. (Luria A. R. Fundamentals of Neuropsychology: textbook. M.: Publishing Center “Academy”, 2003. 384 p., p. 123-126.)
6. Berglund L. and etc. The Reversal Curse: LLMs trained on “A is B” fail to learn “B is A”. <https://arxiv.org/abs/2309.12288>, 26 May 2024.
7. Hagag B. Discover What Every Neuron in the Llama Model Does // Towards Data Science. <https://towardsdatascience.com/-0927524e4807>, Oct 25, 2024.
8. Хомяков А. Б., Чижик П. Новый способ нахождения аналогов как возможность исследования языка, мышления и построения систем искусственного интеллекта // Философские проблемы информационных технологий и киберпространства. 2024. № 1. С. 77-88. <https://doi.org/10.17726/philIT.2024.1.5>. (Khomyakov A. B., Chizhik P. A new way of finding analogues as an opportunity to study language, thinking and build artificial intelligence systems // Philosophical Problems of IT & Cyberspace (PhilIT&C). 2024. № 1. P. 77-88. (In Russ.). <https://doi.org/10.17726/philIT.2024.1.5>.)
9. De La Cruz R. Frank Rosenblatt’s Perceptron, Birth of The Neural Network // Medium. <https://medium.com/@robdelacruz/frank-rosenblatts-perceptron-19fce9d627f>, 1 November 2023.
10. Tiehen J. Perception as controlled hallucination // ResearchGate.org. https://www.researchgate.net/publication/359601789_Perception_as_controlled_hallucination march 2022.
11. Хоффман Д. Как нас обманывают органы чувств. М.: АСТ, 2022. (Hoffman D. How Our Senses Deceive Us. M.: AST, 2022.)

ФИЛОСОФСКИЕ ПРОБЛЕМЫ информационных технологий и киберпространства

№ 2 (26) 2024

ISSN 2305-3763

Свидетельство о регистрации
средства массовой информации
Эл. № ФС77-50786

<https://cyberspace.pgu.ru>

Цели сетевого журнала «Философские проблемы информационных технологий и киберпространства»:

- Повысить статус социо-гуманитарного знания в фундаментальных проблемах развития информационного общества.
- Осветить новые теоретические междисциплинарные направления в современных областях философии техники, социальной эпистемологии, когнитивных науках, теории искусственного интеллекта.
- Расширить сферу профессионального диалога в области информационных технологий, теории киберпространства, виртуалистики, искусственного интеллекта, когнитивных наук.
- Привлечь перспективных специалистов к совместной работе над проектами в гуманитарно-технологической сфере.
- Осуществить информационную поддержку для российских исследователей, работающих в указанных областях.
- Создать коммуникационную платформу для расширения сотрудничества российских и зарубежных профессиональных сообществ.

Задачи сетевого журнала «Философские проблемы информационных технологий и киберпространства»:

- выпуск журнала, осуществление информационной деятельности,
- предоставление полнотекстового доступа к научным статьям,
- продвижение научного контента в отечественных и мировых базах данных,
- менеджмент и аналитика публикационной активности.

Журнал публикует научные статьи, краткие сообщения, обзоры научных мероприятий, рецензии, аннотации. Все материалы, публикуемые в журнале, проходят анонимное рецензирование с рассылкой мотивированных заключений.

Принципиальная политика редакции журнала - открытый доступ (Open Access) к научной информации, бесплатные публикации, постоянная работа над повышением качества научного контента.

В журнале публикуются статьи по следующим областям науки, группам научных специальностей и научным специальностям (согласно Номенклатуре научных специальностей, утвержденной Министерством науки и высшего образования РФ):

1. ЕСТЕСТВЕННЫЕ НАУКИ

1.2. Компьютерные науки и информатика

- 1.2.1. Искусственный интеллект и машинное обучение
- 1.2.2. Математическое моделирование, численные методы и комплексы программ

5. СОЦИАЛЬНЫЕ И ГУМАНИТАРНЫЕ НАУКИ

5.6. Исторические науки

- 5.6.6. История науки и техники

5.7. Философия

- 5.7.1. Онтология и теория познания
- 5.7.2. История философии
- 5.7.6. Философия науки и техники
- 5.7.7. Социальная и политическая философия
- 5.7.8. Философская антропология, философия культуры

5.9. Филология

- 5.9.8. Теоретическая, прикладная и сравнительно-сопоставительная лингвистика
- 5.9.9. Медиакоммуникации и журналистика

РУБРИКИ ЖУРНАЛА:

- Философия языка и компьютерная лингвистика
- История и философия информационных технологий
- Философия когнитивных наук (вычислительные подходы)
- Виртуалистика
- Философия сознания и методология искусственного интеллекта
- Гуманитарное измерение робототехники
- Информационное общество
- Футурология