APPLICATION OF LARGE LANGUAGE MODELS IN DECISION SUPPORT SYSTEMS. PART I: Explanation Models and Large Language Models

A. A. Kulinich

Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

✉ alexkul@rambler.ru

Abstract. Large language models (LLMs) significantly influence many spheres of life: education, creativity, science, and business. This paper considers the use of LLMs to explain alternative solutions obtained by a decision support system under uncertainty. Classical and pragmatic models of explanation proposed by philosophers are discussed. The goals and tasks of explanation in decision support processes under uncertainty are formulated. The operation of LLMs is conceptually analyzed, and their current capabilities in solving typical test tasks are assessed. The main techniques of prompting (a system of queries to a language model) are considered; with these techniques, a language model can be tuned to generate explanations for alternative solutions to particular tasks in a subject area. Finally, prompt techniques for supporting pragmatic and classical theories of explaining alternative solutions are considered.

Keywords: decision support, explanation models, explanation goal, explanation tasks, large language model (LLM), prompt techniques.

INTRODUCTION

To support decision-making in social, political, economic, and organizational systems under uncertainty, "soft" system analysis is used. It is based on the principle of bounded rationality [1]: due to the limited cognitive resources of human model developers, an object's simplified system model can be formed with a hypothetical structure, and its parameters may have linguistic values [2]. Within soft system analysis, model building involves literature examination and expert assessments, i.e., the source data are unstructured data (free text). A model describes an object in a limited natural language proposed by an expert.

Cognitive maps are a well-known mathematical tool of soft system analysis [3, 4]. The results of such a model are presented in an expert's limited natural language and interpreted by an expert in terms of his/her knowledge, which may be insufficient to obtain a new solution. An interpretation should include relevant knowledge of a subject area that, however, lies outside the expert's limited language and a simplified model of the situation. By interpreting the results of modeling, an expert attempts to explain and link the model processes with real-world ones. In a situation model, causal chains of inference can be formally constructed to explain the result [5]. However, these chains of reasoning are expressed in a limited natural language and can ensure the rigor of inference within the accepted assumptions and restrictions, but not the practical applicability of the solution. The real world is much more diverse and richer.

In psychology, decision-making is believed to be carried out by human intelligence in a psychological environment called a mental space. A mental space reflects the knowledge and life experience of an expert gained throughout his/her life. All mental operations—reasoning, generalization, interpretation, explanation, and assessment of possible solutions—take place in a mental space. These operations of decision-making under uncertainty can be supported by artificial intelligence (AI) systems, particularly large language models (LLMs).

Nowadays, both the number and variety of LLMs are growing rapidly. LLMs GPT-2 [6], GPT-3 [7], InstructGPT [8], and GPT-4 [9] from OpenAI, as well as BERT [10], RoBERTa, and ALBERT [11] from Google DeepMind, have become widespread and are intensively used. Developments from Meta AI*, such as LLaMA (Large Language Model Meta AI) [12], are also actively applied. DeepSeek [13], the Chinese language model, is very popular. In Russia, LLMs GigaChat-2 from Sber [14] and Yandex GPT 5.1 from Yandex [15] are widely used. The state-of-the-art advances in the field of LLMs were comprehensively reviewed in [16].

LLMs are applied to solve various tasks in economics, the social sphere, medicine, education, public administration (at the federal, regional, and municipal levels), etc. We mention some reviews of the application of such models: in scientific research [17], for forecasting complex economic systems in modern conditions [18], in public administration [19], in healthcare [20], in organizations and the banking sector [21, 22], as digital assistants [23], and in business analytics and decision-making [24].

A methodology for using LLMs to reference and generate texts in public administration tasks was proposed in [25]. According to the strategic direction in the field of digital transformation of public administration in the Russian Federation [26], it is necessary to introduce AI systems, big data, and the Internet of Things into the sphere of public administration.

Within the Strategic Foresight Session on Fundamental Research in the Field of AI [27], the following priority areas in this field were identified: machine learning architectures and algorithms, computing and data for AI, fundamental and generative models, human–AI interaction, and applied research for science, education, and the social sphere.

This multi-part study is focused on current issues of human–AI interaction in decision-making under uncertainty.

The goal of this study is to develop methods and approaches (models and algorithms) to support decision-making under uncertainty using LLMs.

The main tasks to be solved for achieving this goal are as follows.

First, it is necessary to understand whether the explanatory texts generated by a language model are explanations in the sense of philosophical theories of explanation, which form the fundamental basis of research methodology. To do this, we will analyze the existing explanation models and reveal their core essence.

The second task is to tune an LLM to generate accurate explanations of situations for decision support tasks under uncertainty in tandem with a decision-maker (DM).

* Meta Platforms, Inc. has been recognized as extremist in Russia, and its activities are prohibited within the Russian Federation.

The third task is to assess the quality of a decision support system (DSS) that includes a DM and an LLM as an assistant.

The study consists of two parts. In part I, we describe the first two tasks. In particular, three classes of philosophical explanation models accepted in research methodology are identified. With these classes of explanation models, it is possible to understand whether the text generated by an LLM represents an explanation from a scientific viewpoint and to attribute the text to one of the classes. The role, purpose, and tasks of explanation in decision processes under uncertainty are also defined here. Information on the current performance of LLMs from leading companies is provided. Based on this information, it is possible to judge the applicability of language models as assistants to DMs in decision processes. The possibilities of tuning an LLM using the techniques of prompting (a system of queries to a language model) to generate explanations in terms of the selected classes of explanation models and explanation tasks solved under conditions are determined.

Part I of the study presents the basic concepts and definitions that will be used in part II. The former will also be helpful for those trying to apply an LLM in a DSS for the first time.

Part II of the study will deal with measuring the quality of a cognitive DSS composed of a human (DM) and a language model. To determine the quality of such a system, it is necessary to measure a latent variable, i.e., the DM's satisfaction with an explanation given by a language model. Criteria for assessing the DM's satisfaction degree when solving the explanation tasks outlined in part I of the study will be formulated. A reasonable respondent model will be proposed. An example will be provided where the reasonable respondent model will be used to assess explanations of two Russian LLMs, namely, Sber's GigaChat 2.0 and Yandex GPT 5 Pro.

1. EXPLANATION MODELS IN RESEARCH METHODOLOGY

Philosophical, sociological, and mathematical dictionaries and encyclopedias offer various definitions of an explanation, ranging from its properties and structure to its role in scientific cognition. Here are some generalized definitions.

In the philosophical encyclopedia [28], an explanation is a line of reasoning whose premises contain sufficient information to deduce a description of a phenomenon being explained. An explanation is an answer to the following question: why does this phenomenon occur?

The function of an explanation is defined as that of scientific cognition, the revelation of the essence of an object under study; it is realized by comprehending the law the object obeys, or by establishing the links and relations that determine its essential traits [29].

In research methodology, an explanation is a cognitive procedure aimed at enriching and deepening knowledge of real-world phenomena by incorporating them into a structure of definite links, relations, and dependencies to reveal the essential traits of a given phenomenon [30].

In philosophy and research methodology, the challenge is to create a unified universal theory of explanation applicable in various fields of human activity, e.g., during knowledge acquisition in scientific disciplines such as physics, chemistry, biology, sociology, etc. Variants of the general theory of explanation being created are different explanation models proposed by different researchers or groups of researchers. The explanation models being developed are based on numerous examples of theories of the real world; in general, these models reflect the mechanisms of human intelligence in the processes of research, comprehension, and justification of empirical observations and facts [31].

It is important to understand the extent to which artificial LLMs can explain real objects or events in decision processes under uncertainty.

Note that there are many test programs (benchmarks) to verify and assess the capabilities of language models in the areas of general knowledge, logical reasoning, common sense, etc. It is topical to investigate the capabilities of a language model and assess its quality as an explanation system in decision support processes.

Let us consider the main explanation models proposed in the philosophy of science.

1.1. The Deductive-Nomological Model of Explanation

According to this model, a scientific explanation consists of two main elements: the *explanandum*, which is a statement "describing the phenomenon to be explained," and the *explanans*, which is "statements given to explain the phenomenon" [32]. For explanations (explanans) to explain the explained (explanandum), several conditions must be satisfied.

- The explanandum must be a logical consequence of the explanans, and the statements making up the explanans must be true [32]; i.e., the explanation must take the form of a deductive argument in which the explanandum is derived from the premises making up the explanans. This is the "deductive" component of the deductive-nomological model (DN model).
- The explanans must contain at least one "law of nature," representing a necessary premise for the inference in the sense that if this premise disappears, the inference of the explanandum will be invalid. This is the nomological component of the model [32]. ("Nomological" is a philosophical term that essentially means "lawful, legal") [32].

Formally, the explanation problem for this model (and for all models considered below) has the following most general statement.

An explanation model is a tuple $<\Lambda, Ex1, O, Ex2>$, where:

- Λ is a set of laws; C. Hempel [32] distinguished several classes of laws:
  - ○ universal laws L_c that apply to all areas of knowledge, such as the laws of logic and mathematics;
  - ○ particular laws L_s of separate sciences, e.g., the laws of chemistry, physics, biology, sociology, political science, etc.;
  - ○ individual facts L_f, which obey more general laws.

Thus, in the DN model, the set of laws Λ includes three subsets of the laws defined above, i.e., $\Lambda = \{L_c, L_s, L_f\}$.

The concept of a law is ambiguous (law of nature, state law, etc.), so we will provide several definitions for further reasoning. In philosophy, a law is a necessary connection (interrelation, relationship) between events, phenomena, and also between the internal states of objects, determining their stability, development, stagnation, or destruction. In a philosophical sense, a law is understood to mean objective connections between phenomena and events that exist regardless of whether they are known to anyone or not [33].

A law is a statement, expressed verbally or mathematically, describing objectively existing relationships and connections between various scientific phenomena and objects [28]. A law is proposed as an explanation of facts and is recognized as consistent with them by the scientific community at a certain stage. A law whose validity has been established not from theoretical considerations but from experimental data is called an empirical law.

- $Ex1$ is the explainable, i.e., the situation or phenomenon that needs to be explained.
- $O = \{o_j\}$ is a set of facts that characterize the explainable situation $Ex1$.
- and finally, Ex2 is an explanation.

Definition 1. A DN explanation of a certain phenomenon or situation $Ex1$ is a mapping DN: $L_1(o_1),\ldots, L_k(o_k) \rightarrow Ex2$, where $L_i()$ is a law, $L_i(.)\in\Lambda$; o_j are facts, variables of the law $L_i(.)$, all facts are true, i.e., the probability $P(o_j) = 1$; and finally, DN is a procedure of deductive inference of the explained phenomenon $Ex1$ from the laws L_i and facts o_j. ♦

An explanation that satisfies Definition 1 is called a causal-nomological explanation. This definition implies the application of a deterministic law in the explanation. However, laws from the fields of quantum physics, biology (genetics), as well as sociology and other humanities, are probabilistic in nature. Therefore, the framework of DN explanations defines deductive-statistical and inductive-statistical explanations.

Definition 2. If, in a DN explanation, a general statistical law $L_i^*(.)\in\Lambda$ is used to define explanatory facts, such an explanation is called deductive-statistical (SN explanation). ♦

In this case, the explained object or situation $Ex1$ may include statistical characteristics of a random process (samples, etc.), the mean $M(o_j)$ and the standard deviation $\sigma(o_j)$, and the nomological probability of the explanation $Ex2(M(o_j), \sigma(o_j))$ must be defined.

In this case, the mapping of an explanation from Definition 1 is written as follows:

$$SN: L_1(o_1),\ldots, L_i^*(M(o_j), \sigma(o_j)),\ldots,$$
$$L_k(o_k) \rightarrow Ex2(M(o_j), \sigma(o_j)), L_i^*(.)\in\Lambda.$$

Definition 3. If at least one fact in a DN explanation is obtained by inductive reasoning (the subjective probability of this fact is defined), then the explanation is called inductive-statistical (IN explanation). ♦

It is believed that an inductive fact must have a certain subjective probability. Usually, the probability threshold is set at 0.5, i.e., $o_j| Pr(o_j) > 0.5$.

In this case, the mapping of an explanation from Definition 1 is written as follows:

$$IN: L_1(o_1),\ldots, o_j \mid Pr(o_j) > 0.5,\ldots,$$
$$L_k(o_k)\rightarrow Pr(Ex1), L_i(.)\in\Lambda$$

Unlike deductive-statistical explanations, which are based on a large number of observations of random variables and can be considered a real statistical law, inductive conclusions are formed from a small number of observations and are subjective. In the DN explanation model, inductive facts are not reliable without specifying their context and estimating their probability. Hempel [32] called deductive-statistical and inductive-statistical explanations statistically relevant and inductively relevant, respectively.

In the DN model, explanation validity criteria are proposed. The first is the explanation generalization criterion, and the second is the explanation symmetry criterion. Staying within the mathematical formulations of the DN model, we will define the generalization of explanation elements. In the most general form, generalization is the replacement of an explained concept $Ex1$ and explaining facts $(o_1,\ldots, o_k)$ with the names of the classes or categories they belong to. Let $Kl(Ex1)$ and $Kl(o_j)$ denote the names of the classes of the explained concept and facts, respectively.

Definition 4. A generalization of the DN explanation model is an explanation obtained by replacing the name of the concept $Ex1$ with the name of its class $Kl(Ex1)$ and replacing the names of the facts $(o_1,\ldots, o_k)$ with the names of their classes $(Kl(o_1),\ldots, Kl(o_k))$. ♦

Explanation validity criterion 1: a DN explanation DN: $L_1(o_1),\ldots, L_k(o_k)\rightarrow Ex2$ (see Definition 1) is valid (true) if its generalized DN explanation DN: $Kl(L_1(o_1),\ldots, Kl(L_k(o_k))\rightarrow Kl(Ex2)$ is true. ♦

It is possible to prove the validity of an explanation formally based on generalization if its underlying law refers to the general laws L_c and there is a mathematical model of the object or situation. In this case, there may exist a rigorous mathematical or logical proof of the DN explanation, e.g., in predicate logic, etc. However, in many cases (for instance, in the humanities), such models are absent or too abstract. Then this explanation validity criterion can be confirmed or rejected by an expert, via substituting the names of the classes of concepts of the explained and the facts into the explanation. This criterion allows one to assert the truth and universality of the law chosen for the explanation.

Another explanation validity criterion is the inference symmetry criterion.

Explanation validity criterion 2: A DN explanation DN: $L_1(o_1),\ldots, L_k(o_k)\rightarrow Ex2$ (see Definition 1) is valid (true) if the converse DN explanation DN: $Ex2\rightarrow L_1(o_1),\ldots, L_k(o_k)$ is true. ♦

The existence of a converse explanation inference indicates its consistency. This validity criterion can be formally verified if a known universal law is used for the explanation, and the model of the object or situation is known as well. For this purpose, one can involve the mathematical apparatus of proof theory [34]. In other cases, the consistency of the deductive inference given in an explanation can be verified by an expert.

The DN explanation model is considered a classical model of scientific explanation of objects or situations. However, under the explanation validity criteria, this model turns out to be rather cumbersome and difficult to comprehend. A complete and detailed explanation obtained using the DN model is considered an ideal scientific explanation and is called the "hidden structure" of the explanation.

According to some philosophers, a detailed explanation is excessive and therefore so-called “explanation sketches” are acceptable [32]. By assumption, an incomplete explanation obtained, e.g., by generalizing an ideal explanation, will contain a component that explains the ideal explanation. In this case, the matter concerns explaining an ideal explanation in understandable terms, such as “common sense.” *Common sense* is a set of skills, ways of thinking, and views of the surrounding reality that are developed, used, and shared by people in their everyday practical activities.

An explanation is commonly considered to be something that provides understanding. In this regard, one task of the theory of explanation is to identify the structural features of explanations that provide understanding. Understanding is a universal mental operation associated with the assimilation of new content and its incorporation into a system of established ideas and concepts [35]. It is believed that explanation sketches in commonsense terms can provide an understanding of a phenomenon.

Let common sense be certain empirical laws accepted by society. We add commonsense laws L_{cs} to the set of laws Λ in the DN model: $\Lambda = \{L_c, L_s, L_f, L_{cs}\}$.

Within the hidden structure strategy, an explanation can then be represented as follows.

Definition 5. A hidden structure explanation $Ex2^*$ of a DN explanation $Ex2$ of a certain object or situation is a mapping DN^*: $L_{cs}^1(L_1(o_1)),\ldots, L_{cs}^k(L_k(o_k))\rightarrow Ex2^*$, where $L_{cs}^j()$ are commonsense laws $L_{cs}^j(.)\in\Lambda$; $Ex2^*$ is a commonsense explanation; and finally, DN^* is a deductive inference procedure of the phenomenon being explained from commonsense laws and facts. ♦

A hidden structure explanation can be viewed as an interpretation of an ideal explanation of the DN model. The hidden structure explanation validity can be verified if a homomorphism relation is defined between the basic L_c or particular L_s laws and the commonsense laws L_{cs}, i.e., Φ: $L_c\rightarrow L_{cs}$ or Φ: $L_s\rightarrow L_{cs}$. Under a homomorphism, the elements and relations of the laws L_c, L_s are mapped to those of the commonsense laws L_{cs} while preserving causal relations. Under a homomorphism, the inverse mapping is not unique.

In this case, validity criteria 1 and 2 may fail. Such an explanation will be called empirical.

Philosophers believe that explanations in terms of commonsense laws (explanations based on similarities, analogies, metaphors, etc.) play an important supporting role when formulating an ideal scientific explanation. This fact is due to the ambiguity of the inverse mapping Φ^{-1}: $L_{cs} \rightarrow \{L_{ci}\}$, i.e., a set of rigorous DN explanations based on general laws $\{L_{ci}\}$ can be obtained from an explanation based on the commonsense laws (L_{cs}). This creates the possibility of choosing the best explanation model and stimulates the researcher’s thinking.

1.2. The Unificationist Theory of Explanation

The following explanation model is based on the unificationist theory, which states that scientific explanations can be represented as a uniform description of various objects and phenomena. This theory defines a schematic proposition of the following form [36]: “For all *X*, if *X* is *O* and *A*, then *X* is *P*.” Here, *X* is a variable (the explained); *O*, *A*, and *P*are variables and facts, e.g., some properties of the explained.

The left part of this proposition is the premise ($Prm = (\forall X(O, A))$), and the right part is the inference, i.e., explanation ($\text{Inf} = (X\rightarrow P)$). P. Kitcher introduced the concept of schematic arguments (sequences of schematic propositions) and a classifier to determine which arguments are premises/conclusions/inference rules [36].

Within his approach, a rule for filling (replacing) abstract variables (*X*, *O*, *A*, *P*) with subject variables (concepts from a subject area) is defined. A definition of an argument pattern is given, including arguments, their classification, and filling instructions. A pattern filled with subject variables is considered to be an explanation. Of course, it is possible to obtain many explanation patterns, but which one is correct?

Kitcher introduced the concept of an explanatory reserve [36], i.e., a set of argument patterns containing a set of beliefs shared by scientists at a particular time. To prove that a particular inference—an explanation pattern—is a sound or acceptable explanation, one needs to show its belonging to an explanatory reserve.

The inference procedure in this case is to substitute the values of the subject variable into the argument pattern and check the pattern’s consistency with the explanatory reserve. If consistency holds, then the explanation is obtained; otherwise, the pattern is filled with new data. Strictly speaking, the schematic proposition of this model describes inference in first-order logic. In such logic, a premise explicitly contains a certain law used for inference. For example, in the statement “All humans are mortal” (this is a law), “Socrates is a human” (this is a fact), and “Socrates is mortal” (this is a conclusion), the inference is obtained based on the law.

In the unificationist explanation model, there is no explicit reference to a certain law. A reference to a law in an explanation provides its important property, i.e., the causality of the explanation. This led to philosophical debates about the correctness of such an explanation model. However, according to Kitcher, first, there are examples of explanations without causal relationships, and second, the argument patterns of premises and conclusions should be a generalization of the historical experience of humans and society in various spheres of activity, reflecting their cultural traditions [36]. The long evolutionary process in society forms beliefs, particularly including causal relationships. All these social beliefs are added to the explanatory reserve of the model and are used to validate the explanation pattern. In general, this explanation model suffers from some drawbacks; but, agreeing with its author, we can consider social beliefs to be a certain empirical social law L_{so} and add them to the set of laws Λ of the classical explanation model to get its significant extension with $\Lambda = \{L_c, L_s, L_f, L_{cs}, L_{so}\}$.

Definition 6. An empirical explanation $Ex1^{emp}$ of a certain object or situation is a mapping DN^{emp}: $L_{so}^1(o_1),\ldots, L_{so}^k(o_k)\rightarrow Ex2^{emp}$, where $L_{so}^i()$ is an empirical social law $L_{so}^i(.)\in\Lambda$; o_j are facts, variables of the law $L_{so}^i(.)$, all facts are true, i.e., the probability $P(o_j) = 1$; $Ex2^{emp}$ is an explanation based on empirical laws; and finally, DN^{emp} is a procedure for substituting an empirical explanation $Ex2^{emp}$ from the set of explanatory reserves ER, $Ex2^{emp}\in ER$. ♦

In this case, validity criteria 1 and 2 for an empirical social explanation can be applied by an expert. However, due to the empiricism of a law L_{so}, nothing can be said about the reliability of the explanation (or even its probabilistic assessment). The fulfillment of these criteria will indicate the correct and consistent inference of the explanation, but not its reliability. Various reference books and encyclopedias can serve as an explanatory reserve.

1.3. The Pragmatic Theories of Explanation

The explanation models discussed above neglect the role of humans in the explanatory process. There are pragmatic theories of explanation directly involving humans both in the process of generating explanations and in the process of consuming explanations (as end users).

An explanation theory contains “pragmatic” elements if they require a mandatory reference to facts about the interests, beliefs, or other psychological characteristics of those who give or receive the explanation, and a mandatory reference to the context in which the explanation arises [37].

The authors of the classical DN explanation model agree that pragmatic elements play some role in the process of providing or receiving explanations; but they believe that there is a non-pragmatic core of explanation, and its description is the main task of explanation theory [32].

Contemporary pragmatic theories of explanation are based on constructive empiricism, as outlined by American philosopher B. van Fraassen. According to his viewpoint [37], the goal of science is to construct “empirically adequate” theories yielding true or accurate descriptions of observable phenomena, and research is more about construction than discovery, i.e., building models that must be adequate to the phenomenon rather than discovering the truth related to the unobservable. “Science aims to give us theories that are empirically adequate; and acceptance of a theory involves as belief only that it is empirically adequate.” [37]. Van Fraassen’s “empirical adequacy” means the coincidence of the empirical manifestations of the theoretical model of a phenomenon and the phenomenon itself.

The theory of constructive empiricism complements the theory of scientific realism, which is based on rigorous inference, laws, and proofs. However, as claimed by van Fraassen, the rigor of inference in the theory of scientific realism is based on unproven postulates and axioms, which limits its application in the humanities. While elaborating the theory of constructive empiricism, the author referred to the model of semiotics—the science of signs and sign systems, in which a phenomenon observed is represented at three levels: syntactic, semantic, and pragmatic [38].

A rigorous logical explanation obtained within the theory of scientific realism can be represented as a syntactic abstract structure. In constructive empiricism, such logical structures are interpreted in semantic (meaningful) terms.

In constructive empiricism, a human is an observer of manifestations of reality, a constructor of an empirical model of reality, and its validator. Consequently, “accepting” a theory means only believing in its empirical adequacy [37]. The issues of applying the constructive approach to the study of social systems were considered in [39].

The explanation model proposed within the theory of constructive empiricism is a triple of the form

$$Q = <T_k, X, RL>,$$

where Q is a query; T_k is a topic or context; $X = \{X_1,\ldots, X_n\}$ is a set of contrasting responses, essentially alternative responses (classes of responses); and finally, RL is the relevance relation between the topic and the alternative response.

Let us comment on this model. An explanation is not just a set of judgments but a response to the query: “Why?” It always arises in a definite context and is defined by three factors. The first factor is the topic, i.e., what is being asked about (T_k). The second factor is the contrast class, consisting of a set of statements alternative to the topic. The third factor is the relevance relation between the topic of the query and the contrast class, which defines what can serve as an explanation.

Suppose that for a particular topic T_k and a given set of alternative responses X, a set of pairs—topic and alternative response—can be formed by the direct product $T_k \times X$.

Definition 7. For a pair $(T_k, X_j) \in T_k \times X$, $X_j \in X$, there exists a relevance relation $RL \subseteq T_k \times X$ if the subject believes and is convinced that X_j responds to a query Q in the context T_k. In this case, the explained Q and the explanation X_j are empirically adequate. ♦

A statement X_j is relevant to a query Q if and only if there is a relation RL for the pair (T_k, X_j). A relevance relation cannot be established unambiguously. For example, consider the query Q = "Why does blood circulate in the body?" and two possible responses, X_1 = "Because the contraction of the heart causes blood to move through the arteries" and X_2 = "In order to deliver oxygen to all tissues of the body."

Clearly, there are two alternative responses (two contrasting classes), but it is impossible to establish a relevance relation between the topic of Q and the contrasting class X beyond the context of this query [37].

The query "Why?" identifies a certain problem and sets a definite context. At the same time, the response to it includes a theoretical context—a scientific explanation. For this context of the query, the empirically adequate response is X_1 = "Because the contraction of the heart causes blood to move through the arteries." Here, a causal relevance relation holds. However, if we change the context of the query, Q = "Why does blood circulate in the body?", the empirically adequate response will be X_2 = "In order to deliver oxygen to all tissues of the body," with a functional relevance relation.

It is believed that "…A scientific explanation is not a piece of pure science but an application of science. It is a use of science to satisfy certain of our desires; these desires are always specific to a certain context, but they are always desires for descriptive information. ... The exact content of the desire, and the appraisal of how well it is satisfied, varies from context to context." [37].

Summarizing the explanation models discussed, we emphasize the main differences between pragmatic explanation models and the family of DN models. Recall that the family of DN explanation models (called classical in the literature) was developed within the theory of scientific realism, and such models are characterized by the following properties:

– An explanation is based on laws (deterministic, statistical, or empirical).

– The inference of the explained through observable facts is based on laws and must be correct and true (reliable).

– An explanation reflects the objective laws of nature and/or society and is independent of the psychological characteristics of a subject (his/her interests, beliefs, desires, assessments, etc.).

The pragmatic theories of explanation are based on the theory of constructive empiricism and are characterized by the following properties:

• A subject is included in an explanation model. An explanation is formed considering the subject's psychological characteristics and the context formulated in his/her query.

• A response is formed based on the subjective empirical adequacy of the explanation. The statement regarding the empirical adequacy of an explanation is much weaker than that regarding the truth (reliability) of the explanation in DN explanation models. However, the empirical adequacy of an explanation satisfies the subject's research needs.

• An explanation contains scientifically grounded statements at the level of semantics—the meaning of the explained—rather than at the level of inference in terms accessible to a limited circle of narrow experts, as is done within the theory of scientific realism.

• An explanation is aimed at satisfying the needs and desires of a particular subject in his/her context of interest in order to obtain additional relevant information and expand his/her worldview (mental space).

The above two classes of models form competing theories of explanation, but they can complement each other in decision support processes.

Note that the literature provides explanation models for different subject areas, such as explanation models in sociology, engineering, and mathematics, explanation models of consciousness, etc. Some philosophical explanation models from research methodology have been considered above. Other models of scientific explanation can be found in philosophical literature, e.g., in the review [9]. Here, we briefly overview the well-known, qualitatively different classes of explanation models so that texts generated by an LLM can be expertly assigned one of the classes of scientific explanation. As a result, it becomes possible to assess the quality of the language model's explanations and, therefore, its utility for decision support tasks.

1.4. The Goal of Explanation in Decision Support

The literature on various competing explanation models is quite extensive. Here, researchers interpret the very concept of explanation in different ways, paying little attention to the goals of explanation—what explanations are used for [31]. Obviously, the goal of explanation is determined by the goal of scientific research, which depends on the object of study. The most commonly proposed goals of research are proof, structural analysis, the principle of operation (functioning), development prediction, control of a complex object, etc. The goal of explanation also depends on the field of research; e.g., in the practice of teaching mathematics, the following types of explanation are distinguished: explanation to justify, explanation of a problem solution, explanation to reveal meaning, and explanation to verify.

However, the core of different explanation models in different subject areas is the connection between explanation and understanding. It is believed that explanation provides understanding. Understanding is a universal mental operation associated with the assimilation of new content (new content is included in the explanation) and its inclusion in a system of established ideas and beliefs (in essence, in the subject's knowledge system) [35].

In fact, an explanation is a description of the main properties of an object or situation under study (possibly linking them to known laws) that triggers a universal mental operation—understanding—to integrate this new information into the subject's current knowledge.

According to A.M. Sokhor's connection between understanding and explanation [40], those who understand the explanation constantly and easily move from the real object to its "ideal" model and back can repeat all the cognitive operations of the explainer and understand why these particular operations have been performed in the corresponding sequence. (It is the "ideal" model of an object that is created by abstractions.) Thus, explanation is a bridge linking the real world and the mental processes of human thinking and is intended to form an abstract, idealized mental model of an object or situation. The mental model can then be expressed in a rigorous mathematical language and become a law.

In this case, the goal of explanation can be formulated somewhat differently as applied to decision support. First, an explanation is aimed at forming an information environment of an object or situation, including the parameters of the object or phenomenon, its structure, belonging to a certain class (classification), etc., in the mental space of the subject. In this case, there are various relevance relations of an explanation that meet the subject's needs and the non-rigorous empirical adequacy of the explanation.

Second, an explanation is aimed at forming or identifying causal relations, rigorous inference based on known or newly discovered laws or regularities, for making decisions to control an object or situation.

Thus, the goal of explanation in decision support processes has two components:

– The research component is aimed at creating an information environment for decision-making and developing alternative solutions.

– The practical component is aimed at adopting, justifying, and implementing alternative solutions to control an object or situation.

In this case, competing explanation models—the classical (DN model) and pragmatic ones—complement each other. When treating decision-making as a research process, we can identify two stages of this process. The first stage—research—is a series of sequential queries and responses/explanations to form the information environment necessary for elaborating alternative solutions. At this stage, pragmatic explanation models are appropriate, as they yield empirically adequate explanations of alternative solutions without restricting the researcher's desires and needs.

The second stage is to select the best alternative using the DN explanation model, which justifies alternatives based on known laws and rigorous inference.

2. LARGE LANGUAGE MODELS AND ASSESSMENT OF THEIR CAPABILITIES

LLMs are deep learning neural networks trained on huge arrays of text data. They are based on the transformer architecture, which includes a set of neural networks consisting of an encoder and a decoder. Transformer is a neural network architecture designed for natural language processing and many other machine learning tasks [41]. An encoder is an element of the transformer architecture that converts a text corpus into vector representations: each word is described by a vector of probabilities of its joint use with other words in the text corpus, and the information about the structure and relationships between words is preserved. A decoder uses the encoded information to generate a response or make predictions. It "decodes" the received data, creating new text sequences considering previous words, context, and the probabilities of the joint use of words in that context.

The neural network of an LLM is a multilayer structure where each layer consists of a set of artificial neurons connected to neurons of neighboring layers. The number of layers in different LLMs varies and can reach several dozen or more. The neural network is trained with a large text corpus to form a dictionary of tokens, i.e., individual letters or their groups, words, phrases, and sentences with a defined probability of their joint use in a given context. When encoding input information and transferring it between layers, an attention mechanism is used to reveal important elements of the language structure. Attention mechanisms can identify different types of syntactic relations between words, actually separating the syntactic structure of a sentence [42]. During training, a vector space of contextual token vectors is formed, and the (cosine) measure of proximity is determined for tokens. When decoding, tokens (words, phrases, and sentences) close to the query's topic are selected from the vector space to form sentences responding to this query. Selecting tokens from the vector space that are close to the query tokens allows generating a response from the words that frequently occur in the context of this query.

Thus, an LLM produces a statistical response to a query without working at the semantic level or analyzing the meaning of the query and response. The meaning of a response depends on the quality of a corresponding text corpus. If the corpus includes texts from verified sources, the response will be meaningful and understandable.

If a language model does not understand a query or has been trained with incomplete or incorrect data, it tries to guess the response based on existing syntactic patterns, which may lead to false responses. This behavior of an LLM is called hallucination. According to the hallucination phenomenon, LLMs generate factually incorrect or fictitious data that are not based on real information. This behavior is also characteristic of people when, under deficient information or time constraints for problem analysis, they form random or fictitious responses that are far from reality.

The problem of hallucinations in LLMs is crucial in applications with critical requirements for reliable information. Currently, various metrics have been introduced to assess the tendency of language models to hallucinate. Also, several methods for reducing language model hallucinations have been developed: the creation of correct text corpora (manual verification and text cleaning); the formation of correct queries to a language model, prompting the correct response; retraining a language model; and searching for accurate and relevant information in external sources.

Note that language model hallucinations in research and creative tasks can stimulate human intuition and lead to original solutions/decisions. Recall that heuristic methods for solving creative tasks (such as brainstorming and synectics) do not reject absurd and counterproductive alternative solutions as potential intuition impulsion to generate new and original solutions.

Despite this drawback, language models are useful owing to their capability to solve many natural language-written tasks that could previously be settled only by humans.

LLMs are assessed by special programs (benchmarks), which measure their basic qualities: knowledge volume, response accuracy, reliability, etc.

A benchmark uses definite datasets, metrics, and assessment tasks to test a language model, allowing one to compare different models and measure their accuracy. There are benchmarks to test knowledge, logical thinking, reading comprehension, common sense, etc. [43]. Many benchmarks have been developed; let us consider some of them.

Knowledge benchmarks test models in various fields. They assess how effectively a model can recall information from different areas, such as physics, geography, etc. MMLU (*Measuring Massive Multitask Language Understanding*) is a well-known benchmark created to check the model's factual knowledge on various topics, such as the humanities, social sciences, history, computer science, and even law. It includes 57 queries and 15 000 tasks designed to verify the high capabilities of a language model. On this benchmark, GPT-4-omni correctly responded to 88.7% of the queries asked.

Logical thinking benchmarks test the model's capability to "think" step by step and make logical conclusions (inference).

Benchmarks for assessing the mathematical capabilities of language models are used as well. For example, the GSM8K test consists of 8500 middle-school math tasks. Solving them requires the model to perform several steps of elementary calculations. Language models pre-trained for mathematical reasoning perform well on this benchmark; for example, GPT-4 achieves an accuracy of 96.5%.

GPQA (*Graduate-Level Google-Proof Q&A Benchmark*) assesses the logical thinking of a language model using a dataset of only 448 queries. This difficult test, developed by experts in biology, physics, and chemistry, was passed by GPT-4-omni with an accuracy of 53.6% only, while graduate students achieve 65%.

Reasoning language models are being developed by many leading companies. In 2024, OpenAI released a new language model, OpenAI-o1, which demonstrates excellent results in complex reasoning, outperforming humans in tests in mathematics, coding, and natural sciences. In the qualifying round for *the International Mathematical Olympiad* (IMO), this model succeeded in solving 83% of the tasks, while its predecessor (GPT-4o) gave only 13% correct responses. According to the developers from OpenAI, when solving complex test tasks in physics, chemistry, and biology, the model demonstrates results comparable to those of postgraduates.

Reading comprehension benchmarks test the model's capability to interpret natural language and generate appropriate responses. The test is to respond to queries about texts, which allows assessing comprehension and the capability to make conclusions, grasp, and remember important details.

DROP (*Discrete Reasoning Over Paragraphs*) is a benchmark for testing reading comprehension. It challenges models to reason based on their analysis of paragraph content. This benchmark includes 96 000 queries to test the reasoning capabilities of a language model. DROP's queries contain information that requires models to perform mathematical operations such as addition, subtraction, and comparison based on information scattered throughout the text. When responding to these complex queries, GPT-4 achieved an accuracy of 80%, while humans give 96% correct responses on the DROP dataset.

Commonsense benchmarks assess the model's capability to generalize knowledge of the world. Such test sets typically include queries that require extensive encyclopedic knowledge to respond correctly. Commonsense testing in language models assesses the model's capability to make judgments and conclusions consistent with human thinking. Humans form a holistic view of the world through practical experience, while language models are trained on huge datasets without understanding context.

HellaSwag (*Harder Endings, Longer Contexts, and Low-Shot Activities for Situations with Adversarial Generations*) is a benchmark for checking the model's capability to predict a plausible continuation of a given scenario. According to HellaSwag testing results, modern models such as GPT-4 have achieved near-human levels of accuracy.

IFEval (*Inference and Fidelity Evaluation*) is a benchmark for assessing both the accuracy and quality of the generated text. First, a model is assessed by the Inference metric—the capability to generate text on a large volume of data. Then, the quality (Fidelity) of the generated text is assessed. The assessment includes checking the compliance of the generated text with the expected result and assessing the degree of preserving the meaning and structure of the text. Next, the final IFEval score is calculated, which reflects both the model's capability to generate text and the quality of that text. The higher the IFEval score is, the better the model will succeed in text generation.

The **HumanEval benchmark** is a reference dataset designed for objectively assessing the quality of code generated by AI models based on a text description of the task. The benchmark consists of 164 programming tasks manually written for this dataset to ensure their absence in the model's training samples. All tasks are formulated in Python and presented as code snippets with descriptions.

Table 1 presents performance estimates for LLMs from leading developers: GigaChat 2 MAX from the Russian company Sber; Qwen 2.5 72B from leading e-commerce platform Alibaba; Llama 3.3 70B developed by Meta AI; GPT-4o developed by OpenAI; DeepSeek-V3 from Chinese company DeepSeek, owned by the High-Flyer fund; and Yandex GPT5 Lite Instruct from Russian company Yandex. Testing was conducted on benchmarks in the following categories: general knowledge, mathematics, code work, and text generation quality. The data are current [44, 45] as of early 2025. The figures in the table show the percentage of tasks solved.

Testing the capabilities of language models using benchmarks allows objectively assessing and comparing the quality of models from different suppliers and understanding the current level and dynamics of language model development. Language model developers believe that the capabilities of these models are approaching those of humans; hence, they can assist in decision support systems.

However, practical tasks, such as making decisions in complex economic, political, or social situations, require decision-makers to have a comprehensive combination of all possible human intelligence capabilities. Therefore, automated metrics alone cannot cover the entire spectrum of language model assessment, especially when it comes to the subjective aspects of language comprehension and generation. Here, human assessment is much more accurate.

In this case, it is advisable to involve experts or a group of experts to provide an accurate and reliable assessment of the capabilities of LLMs [43], e.g., based on a survey.

Some problems arise in the expert assessment of language model explanations. Here, it is necessary to assess the interaction between a human and an LLM. A language model gives a statistically plausible explanation without understanding its meaning, while a human tries to understand the explanation and integrate it into his/her knowledge system. A human understands a language model's explanation depending on his/her level of knowledge; and therefore, different people, receiving the same response from a language model, may give it different assessments.

In this case, it is required to develop a method for assessing the individual's (non-group) satisfaction with the explanations of a decision situation that he/she receives from an LLM. The solution to this task will be discussed in part II of the study.

Table 1

Performance of LLMs from leading developers

Category	Benchmark name	GigaChat 2 MAX	Qwen 2.5 72B	Llama 3.3 70B	GPT-4o	DeepSeek-V3	Yandex GPT5 Lite Instruct
General knowledge	MMLU (RU)	80.46	78.30	65.08	80.00	73.74	70
	MMLU (EN)	86.00	83.85	78.57	88.7	85.24	75.8
Mathematics	GSM8K	95.68	95.07	92.87	95.00	94.99	87.9
	MATH	77.26	78.74	62.80	76.60	85.48	82.0
Code work	HumanEval	87.20	86.60	86.0	84.00	91.46	71.8
Quality of text generation	IFEVAL (RU)	83.62	84.27	75.12	80.24	84.37	76.9
	IFEVAL (EN)	89.99	90.43	90.83	88.51	92.21	72.6

3. EXPLANATION MODELS AND LARGE LANGUAGE MODELS

The explanation models discussed above form two classes of models: classical (deductive-nomological) models, based on rigorous inference and known objective laws or regularities, and pragmatic models, which consider the goals and desires of the researcher or explainer and the context of the query to be explained.

In these models, a human formulates an explanation based on observable facts and knowledge. Obviously, the knowledge of a human researcher necessary to formulate a response-explanation will be different in each class of explanation models. Benchmarks show the good capabilities of language models in different areas. Can a language model provide an explanation satisfiable for a DM?

The knowledge of a language model is determined by the content of the text corpus used to train it. Therefore, a language model will operate in the class of classical explanation models if it has been trained in the known laws of logic, mathematics, physics, chemistry, biology, etc. In this case, the model is likely to generate rigorous logical explanations based on known laws. The reliability of such inference can be verified, and the logical conclusions and explanations themselves can be used to control a situation.

A language model will operate in the class of pragmatic explanation models if it has been trained in general knowledge, commonsense laws, traditions, etc. According to the theory of constructive empiricism, the resulting empirically adequate explanations are based on the researcher’s beliefs. Such explanations are useful for expanding the mental space of a DM, increasing his/her awareness, and stimulating his/her intuition to generate non-trivial creative decisions.

Training language models to obtain classical explanations or pragmatic explanations from scratch is a very costly procedure. First, it involves collecting, preparing, and cleaning a very large training text corpus relevant to a given task. As a rule, text corpora are prepared by special outsourced teams. The dictionaries of modern language models contain billions of tokens. As is believed, only very large language models exhibit emergent properties, such as the capability to reason. Second, training a neural network requires very high computing power—supercomputers—and, accordingly, significant time and energy. Training a language model from scratch is only affordable for large companies, which allow other developers to use trained models in their tasks.

Pre-trained language models that are publicly available can be retrained with particular texts, focusing them on the solution to particular tasks. In this case, the stages of preparing the text corpus and training the model are also present, but at a lower cost.

An LLM is a huge neural network with a huge number of connections between tokens. Each query to a network activates a definite chain of tokens, forming a response. Also, there are a huge number of options for activating token chains, and therefore, the number of plausible alternative responses is also large.

Currently, *Prompt Engineering* (PE) is being actively developed as an alternative tuning method for an LLM for particular tasks using prompts [46, 47]. A prompt is a hint or sequence of hints for a language model. Prompts are represented as an algorithm or scenario for solving a user task; they are created by the user and can be embedded in a language model. Thus, prompts control the operation of a language model to obtain the best result.

3.1. Techniques of Prompting and Explanation Models

Consider some of the most popular techniques of prompting.

Zero-shot. This technique of prompting implies no input or output examples to retrain a language model. It is a simple query to a language model, and the latter can respond by providing information that does not require multi-step instructions.

Few-shot. This technique of prompting involves examples: several examples of input and output data are used to retrain a language model. With these examples, a language model learns to generate queries following the patterns. For example, a model can be retrained to detect the tone (emotional "color") of texts accurately. Few-shot prompting can be employed as a technique for the contextual retraining of a language model. Examples specify the context of an expected response.

Role-based. This technique of prompting is to assign a role to a language model when it responds to a given query. For example, a model is asked to respond to a query from the standpoint of an economist, financier, lawyer, etc. The query is the same, but the responses will be generated in the contexts of different subject areas [48].

Chain-Of-Thought (CoT). This technique of prompting forces a language model to think step by step. In this technique, a user gives a language model an algorithm for solving a task. It makes the model's thinking closer to human thinking. As a rule, a complex task is decomposed into subtasks, and their sequential solution makes the overall solution more accurate. The application of this technique in LLMs significantly improves the accuracy of solving mathematical and logical problems of the GSM8K benchmark.

According to the authors, CoT leads to higher-quality solutions to linguistic tasks as well as general knowledge and commonsense tasks. This popular technique of prompting, in various modifications, is widely used in reasoning language models [49].

Chain-of-Verification (CoV). It complements CoT by forcing a model to verify all previous steps before taking the next step, making reasoning language models more reliable [50].

Chain-of-Note (CoN). With this technique of prompting, a language model is forced to make "notes" in the process of solving a task, thereby explaining each successive step. As a result, the hallucinations of a language model can be detected [51].

Chain-of-Knowledge (CoK). In this technique, a prompt necessarily includes verified knowledge of a subject area, and a language model can use it to solve a given task. Unlike CoT, a language model relies on known facts, laws, or regularities provided in a user's prompt and then builds a logical chain from them, leading to a specific and reasonable answer. For example, a user includes the laws of physics in the prompt to solve a physics problem, and a language model will generate the solution using these laws. The correctness of the response will be higher than in (simple) CoT [52].

Tree of Thoughts (ToT) [53]. This technique is used to solve complex research tasks or strategic planning tasks when conventional or simple methods of prompting are insufficient. ToT was proposed in [53] as a generalization of CoT. In this technique, attention is paid to the study of responses, which serve as intermediate steps for solving tasks by language models. ToT supports the construction of a response tree (tree of thoughts) whose nodes contain intermediate response texts. Then search algorithms, such as breadth-first search and depth-first search, are applied to the tree to obtain a solution and analyze it.

ReAct Prompting. In this technique of prompting [54], a language model is used in interactive mode with a human to generate chains of reasoning in order to form relevant actions for a given task. With generated chains of reasoning, a language model is able to create, track, and update action plans, as well as to handle error situations. Actions implement interaction with external information sources, e.g., knowledge bases.

ReAct allows a language model to interact with external sources to obtain additional information, resulting in more reliable and accurate responses. ReAct improves the interpretability and reliability of a language model. According to the authors, the best approach is to apply ReAct in combination with CoT; in this case, both internal knowledge and external information obtained by reasoning are used.

Commonsense prompts. Currently, prompts for supporting deductive reasoning based on commonsense knowledge are of great research interest. The fact is that LLMs are trained on large text corpora, and when answering queries, they give statistically better responses. Such responses are correct, but they usually have no elements of deductive reasoning (obvious to humans) and are represented as implicit knowledge. Prompts are developed to obtain a language model's response containing an explanation with deductive reasoning; they are hints for a language model to extract a meaningful answer. Such hints for a language model are generated using external commonsense knowledge bases. These are static knowledge bases ConceptNet [55] and ATOMIC [56], as well as dynamic commonsense knowledge base COMeT [54], in which commonsense knowledge is formed from the context of a query to a language model and is represented in the form of a reasoning graph. Experiments with this dynamic knowledge base [57] have shown an increase in the performance of language models on commonsense benchmarks. In the paper [58], a method was described to generate knowledge from a language model itself and provide this knowledge as additional input data (hints) when answering a query. This method does not require access to a structured commonsense knowledge base but improves the performance of modern LLMs. The issues of generating commonsense responses by a language model based on internal dialog with a language model were considered in [59]. The corresponding method involves no external knowledge bases, but the performance of a language model on commonsense benchmarks with such a prompt has demonstrated an increase. However, according to the authors of [59], the disadvantage of LLMs is the lack of introspection, i.e., knowledge of their knowledge.

Only some techniques of prompting have been discussed above. Currently, many developers offer custom prompts that can be embedded in pre-trained language models and do not require additional training in a subject area. This seems to be an inexpensive option for tuning a language model to solve user tasks.

In decision support, there are heuristic techniques of analyzing situations under uncertainty, which are methods for finding decisions in complex situations. For example, they include Five Whys, aimed at identifying the causes of a problem; the Ishikawa (fishbone) diagram, intended to build a hierarchical model of cause-and-effect relations; the garland and association method, serving to generate creative decisions; SWOT analysis, allowing one to determine the development strategy of an organization, etc. All these techniques have been tested in decision practice, and their algorithms can be considered a prompting algorithm for an LLM.

A prompt can be represented as a control program that extracts the information necessary for decision-making from a language model via sequential steps.

The above techniques of prompting can help a language model implement the explanation models. Let us consider this issue in detail.

For pragmatic explanation theories focused on satisfying the user's desires and interests, suitable techniques are those organizing a dialog with a language model to expand the user's mental space and awareness. Recall that in pragmatic explanation models, there shall exist a relevance relation between the topic of a query and the language model's explanation, and the DM shall be convinced that the explanation is empirically adequate.

The techniques of prompting from Table 2 can be applied to support pragmatic explanation models.

Note that Few-shot allows implementing the unificationist explanation model [36]. Recall that in this model, a pattern is defined, and an explanation selected from an explanatory resource is substituted into this pattern. An explanatory resource is an opinion of prominent scientists, customs, traditions, and information from reference books, encyclopedias, etc.

For classical explanation theories (DN models [32]**)**, aimed at obtaining and justifying an explanation in the form of decision inference based on known laws and regularities, suitable techniques of prompting are provided in Table 3.

We have mentioned two goals of explanation in decision support. These are the research goal, which is achieved by using a language model with prompts implementing pragmatic explanations to increase expert awareness, and the practical goal, which is achieved by using a language model with prompts for classical explanation theories that help obtain and justify rigorous decision inference and its application to control a situation.

The application of different combinations of the existing techniques of prompting or the development of a custom prompt allows an LLM to be tuned for explaining alternative solutions in both the classical and pragmatic theories of explanation.

The issues related to the development and testing of user prompts for solving research and practical decision support tasks are not addressed in this work.

Table 2

The role of prompting in the pragmatic explanation model

The technique of prompting	The role of prompting
Zero-shot	Allows getting a response to any query
Few-shot	Allows getting a response in the form of a user-defined pattern
Role based	Allows getting responses to the same query in different contexts
Chain-Of-Thought (CoT)	Allows representing a complex task as a series of related subtasks and solving them sequentially
Chain-of-Verification	Extends CoT; allows verifying the reasoning of a language model and detecting hallucinations
Chain-of-Note	Extends CoT; allows improving the reasoning quality of a language model by supplementing explanations
Tree of Thoughts (ToT)	Allows building a tree of reasoning instead of a chain of thoughts in CoT; as a result, allows finding a better alternative explanation
ReAct Prompting	Enables an interactive mode between a user and a language model. The user can utilize external data to solve research and planning tasks

Table 3

The role of prompting in the classical explanation model

The technique of prompting	The role of prompting
Chain-of-Knowledge	Allows retraining a language model with a knowledge model that includes a description of the laws or patterns of a particular subject area. As a result, allows increasing the inference reliability of a language model in the subject area under study.
ReAct Prompting	Organizes an interactive mode of operation between a user and a language model with access to external sources of verified data. As a result, allows the use of descriptions of laws from external data sources, which can increase inference reliability
Commonsense	Allows getting a "sketch" of a classical explanation in commonsense terms. Such an explanation has been called a hidden-structure explanation of a deductive-nomological explanation (see above).

CONCLUSIONS

We have considered the application of LLMs for generating and explaining alternative solutions obtained by a decision support system under uncertainty. We have discussed classical (deductive-nomological) and pragmatic models of explanation proposed by philosophers. The goals and tasks of explanation in decision support processes under uncertainty have been formulated. The goal of explanation in a decision support system is to form an information environment for decision-making; it reduces to solving a research problem to formulate alternative solutions and a practical problem to justify and implement the best alternative. The operation of LLMs has been conceptually analyzed, and their current capabilities in solving typical test tasks have been assessed. The main techniques of prompting (a system of queries to a language model) have been considered; with these techniques, a language model can be tuned to generate explanations for alternative solutions to research and practical tasks of decision support under uncertainty.

In part I of the study, we have formulated the basic concepts and definitions that will be used in part II. In turn, part II of the study will be devoted to issues of measuring and assessing the satisfaction of decision-makers with explanations of LLMs. Satisfaction assessments will be conducted for research and practical tasks of explanation under uncertainty. The explanations of two Russian LLMs will be analyzed and assigned to the above classes of explanation models.

REFERENCES

1. Simon, H.A., Rationality as Process and as Product of Thought, in *Decision Making: Descriptive, Normative, and Prescriptive Interactions*, Bell, D.E., Raiffa, H., and Tversky, A., Eds., Cambridge: Cambridge University Press, 1988, pp. 58–77.
2. Checkland, P.B., *Systems Thinking, Systems Practice*, New York: Wiley, 1981.
3. Axelrod, R., *The Structure of Decision: Cognitive Maps of Political Elites*, Princeton: Princeton University Press, 1976.
4. Kosko, B., *Fuzzy Thinking: T*he New Science of Fuzzy Logic, New York: Hyperion, 1993.
5. Kulinich, A.A., Cognitive Maps Verification Based on Processes Explanation, *Large-Scale System Control*, 2010, no. 30.1, pp. 453–469. (In Russian.)
6. *Neural Network GPT-2 by OpenAI. Fast Start*. URL: https://habr.com/ru/articles/440564/ (Accessed February 16, 2026; in Russian.)
7. Brown, T., Mann, B., Ryder, N., et al., Language Models Are Few-Shot Learners, *arXiv:2005.14165*, 2020. DOI: https://doi.org/10.48550/arXiv.2005.14165
8. Ouyang, L., Wu, J., Jiang, X., et al., Training Language Models to Follow Instructions with Human Feedback, *arxiv.org/abs/2203.02155*, 2022. DOI: https://doi.org/10.48550/arXiv.2203.02155
9. Brown, A., GPT-4 Is OpenAI's Most Advanced System, Producing Safer and More Useful Responses, *International Journal of Architectural Computing*, 2024, vol. 22, no. 3, pp. 275–276. DOI:10.1177/14780771241280148
10. Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K., Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding, *arXiv:1810.04805*, 2018. DOI: https://doi.org/10.48550/arXiv.1810.04805
11. Lan, Z., Chen, M., Goodman, S., et al., Albert: A Lite Bert for Self-Supervised Learning of Language Representations, *arXiv:1909.11942*, 2019. DOI: https://doi.org/10.48550/arXiv.1909.11942
12. Touvron, H., Lavril, T., Izacard, G., et al., Llama: Open and Efficient Foundation Language Models, *arXiv:2302.13971*, 2023. DOI: https://doi.org/10.48550/arXiv.2302.13971
13. Bi, X., Chen, D., Chen, G., et al., Deepseek LLM: Scaling Open-Source Language Models with Longtermism, *arXiv:2401.02954*, 2024. DOI: https://doi.org/10.48550/arXiv.2401.02954
14. *GigaChat-2*. URL: https://giga.chat/ (Accessed February 14, 2026; in Russian.)
15. *Yandex GPT 5.1*. URL: https://ya.ru/ai/gpt (Accessed February 14, 2026; in Russian.)
16. Vizil'ter, Yu.V., Priority Research Areas and Key Trends in the Development of AI Technologies, *Trudy 22-oi Natsional'noi konferentsii po iskusstvennomu intellektu s mezhdunarodnym uchastiem* (Proceedings of the 22nd Russian Conference on Artificial Intelligence with International Participation (RCAI-2025)), St. Petersburg, 2025, vol. 1, pp. 7–25. DOI: 10.15622/rcai.2025.001 (In Russian.)

17. Bragin, A., Bakhtizin, A., and Makarov, V., Large Fourth-Generation Language Models as a New Tool in Scientific Research, *Artificial Societies*, 2023, vol. 18, no. 1. DOI: 10.18254/S207751800025046-9 (In Russian.)
18. Bakhtizin, A.R., The Challenges of Forecasting under Current Conditions, *The Economic Revival of Russia*, 2023, no. 2 (76), pp. 53–62. (In Russian.)
19. Kommersant Learned about the Testing of Russian Language Models for Gosuslugi, *Forbes*, February 2, 2024. URL: https://www.forbes.ru/tekhnologii/505447-kommersant-uznal-o-testirovanii-rossijskih-azykovyh-modelej-dla-gosuslug?ysclid=lzfs8wut7z80399139 (Accessed February 14, 2026; in Russian.)
20. Gusev, A., A Review of Russian Artificial Intelligence Systems for Healthcare. URL: https://webiomed.ru/blog/obzor-rossiiskikh-sistem-iskusstvennogo-intellekta-dlia-zdravookhraneniia/ (Accessed February 14, 2026; in Russian.)
21. Shrestha, Y.R., Ben-Menahem, S., and von Krogh, G., Organizational Decision Making Structures in the Age of Artificial Intelligence, *California Management Review*, 2019, vol. 61, no. 4, pp. 66–83.
22. Gasanov, E., Decision Intelligence: Artificial Intelligence with a Human Face, *IT World*, March 22, 2022. URL: https://www.it–world.ru/cionews/1q746pou69z40kokwskckk4gkg4c0og.html (Accessed February 14, 2026; in Russian.)
23. Virtual Assistants, *TAdviser*, August 6, 2022. URL: https://www.tadviser.ru/index.php (Accessed February 15, 2026; in Russian.)
24. AI in Analytics: What is Beyond BI?, *TAdviser*, August 4, 2022. URL: https://www.tadviser.ru/index.php (Accessed February 15, 2026; in Russian.)]
25. Dudikhin, V.V. and Kondrashov, P.E., Methodology of Using Large Language Models to Solve Tasks of State and Municipal Government for Intelligent Abstracting and Automatic Generation of Text Content, *Public Administration*, 2024, no. 105, pp. 169–179. (In Russian.)
26. *Strategic Direction in the Field of Digital Transformation of Public Administration*. Approved by Decree no. 2998-r of the Government of the Russian Federation on October 22, 2021. (In Russian.)
27. Dmitry Chernyshenko Held a Strategic Foresight Session on Fundamental Research in the Field of Artificial Intelligence, *Official Website of the Government of the Russian Federation*, May 30, 2024. URL: http://government.ru/news/51726/ (Accessed February 16, 2026; in Russian.)
28. *Filosofiya: Entsiklopedicheskii slovar'* (Philosophy: Encyclopedic Dictionary), Ivin, A.A., Ed., Moscow: Gardariki, 2004. (In Russian.)
29. *Filosofskii entsiklopedicheskii slovar'* (Philosophical Encyclopedic Dictionary), Il'ichev, L.F., Fedoseev, P.N., Kovalev, S.M., and Panov, V.G., Eds., Moscow: Sovetskaya Entsiklopediya, 1983. (In Russian.)
30. *Novaya filosofskaya entsiklopediya* (New Philosophical Encyclopedia), Stepin, V.S., Ed., Moscow: Mysl', 2001. (In Russian.)
31. Woodward, J., Scientific Explanation, in *The Stanford Encyclopedia of Philosophy*, Zalta, E.N., Ed., Stanford: Stanford University, 2017. URL: https://plato.stanford.edu/archives/fall2023/entries/scientific-explanation/ (Accessed February 15, 2026.)
32. Hempel, C., *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: Free Press, 1965.
33. Regularity (law), in *Bol'shaya rossiiskaya entsiklopediya* (The Great Russian Encyclopedia), Osipov, Yu.S., Ed., Moscow: Great Russian Encyclopedia, 2004–2017. (In Russian.)
34. Dragalin, A.G., *Mathematical Intuitionism. Introduction to Proof Theory*, Translated by E. Mendelson, Providence, Rhode Island: AMS, 1988.
35. Understanding, in *Slovar' po logike* (Dictionary of Logic), Ivin, A.A. and Nikiforov, A.L., Eds., Moscow: Vlados, 1997. (In Russian.)
36. Kitcher, P., Explanatory Unification and the Causal Structure of the World, in *Scientific Explanation*, Kitcher, P. and Salmon, W., Eds., Minneapolis: University of Minnesota Press, 1989, pp. 410–505.
37. Van Fraassen, B., *The Scientific Image*, Oxford: Oxford University Press, 1980.
38. Morris, C., Foundations of the Theory of Signs, in *Writings on the General Theory of Signs*, Sebeok, T., Ed., The Hague: Mouton, 1971.
39. Plotinskii, Yu.M., *Modeli sotsial'nykh protsessov* (Models of Social Processes), Moscow: Logos, 2001. (In Russian.)
40. Sokhor, A.M., *Ob"yasnenie v protsesse obucheniya: Elementy didakticheskoi kontseptsii* (Explanation in the Learning Process: Elements of the Didactic Concept), Moscow: Pedagogika, 1988. (In Russian.)
41. Vaswani, A., Shazeer, N., Parmar, N., et al., Attention is All You Need, *arXiv:1706.03762*, 2017. DOI: https://doi.org/10.48550/arXiv.1706.03762
42. Marecek, D. and Rosa, R., Extracting Syntactic Trees from Transformer Encoder Self-attentions, *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, Brussels, Belgium, 2018, pp. 347–349. URL: https://aclanthology.org/W18–5444.pdf/ (Accessed February 15, 2026.)
43. Barskaya, I., Benchmarks for LLMs, *Unite.AI*, August 28, 2024. URL: https://www.unite.ai/benchmarks-for-llms/ (Accessed February 15, 2026.)
44. *Sber Presented GigaChat 2.0, the New Generation of Its Neural Network*. URL: https://t-j.ru/news/sber-gigachat-2/ (Accessed February 15, 2026; in Russian.)
45. *YandexGPT-5-Lite-8B-instruct*. URL: https://huggingface.co/yandex/YandexGPT-5-Lite-8B-instruct (Accessed February 15, 2026.)
46. Sahoo, P., Singh, A.K., Saha, S., et al., A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications, *arXiv:2402.07927*, 2024. DOI: https://doi.org/10.48550/arXiv.2402.07927
47. *Prompt Engineering Guide*. URL: https://www.promptingguide.ai/ (Accessed December 16, 2023.)
48. Wang, Z., Peng, Z., Que, H., et al., RoleLLM: Benchmarking, Eliciting, and Enhancing Role-Playing Abilities of Large Language Models, *arXiv:2310.00746*, 2024. DOI: https://doi.org/10.48550/arXiv.2310.00746
49. Wei, J., Wang, X., Schuurmans, D., et al., Chain of Thought Prompting Elicits Reasoning in Large Language Models, *arXiv:2201.11903*, 2022. DOI: https://doi.org/10.48550/arXiv.2201.11903
50. Dhuliawala, S., Komeili, M., Xu, J., et al., Chain-of-Verification Reduces Hallucination in Large Language Models, *arXiv:2309.11495*, 2023. DOI: https://doi.org/10.48550/arXiv.2309.11495
51. Yu, W., Zhang, H., Pan, X., et al., Chain-of-Note: Enhancing Robustness in Retrieval-Augmented Language Models,

arXiv:2311.09210, 2023. DOI: https://doi.org/10.48550/arXiv.2311.09210
52. Liu, J., Lui, A., Lu, X., et al., Generated Knowledge Prompting for Commonsense Reasoning, *arXiv:2110.08387*, 2021. DOI: https://doi.org/10.48550/arXiv.2110.08387
53. Yao, S., Yu, D., Zhao, J., et al., Tree of Thoughts: Deliberate Problem Solving with Large Language Models, *arXiv:2305.10601*, 2023. DOI: https://doi.org/10.48550/arXiv.2305.10601
54. Yao, S., Zhao, J., Yu, D., et al., React: Synergizing Reasoning and Acting in Language Models, *arXiv:2210.03629*, 2023. DOI: https://doi.org/10.48550/arXiv.2210.03629
55. Speer, R., Chin, J., and Havasi, C., Conceptnet 5.5: An Open Multilingual Graph of General Knowledge, *arxiv.org/abs/1612.03975*, 2017. DOI: https://doi.org/10.48550/arXiv.1612.03975
56. Sap, M., Le Bras, R., Allaway, E., et al., Atomic: An Atlas of Machine Commonsense for If-Then Reasoning, *Proceedings of the AAAI Conference on Artificial Intelligence*, Honolulu, 2019, vol. 33, pp. 3027–3035.
57. Bosselut, A., Rashkin, H., Sap, M., et al., COMET: Commonsense Transformers for Automatic Knowledge Graph Construction, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, 2019, pp. 4762–4779.
58. Bosselut, A., Le Bras, R., and Choi, Y., Dynamic Neuro-Symbolic Knowledge Graph Construction for Zero-Shot Commonsense Question Answering, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vancouver, Canada, 2021, pp. 4923–4931.
59. Shwartz, V., West, P., Le Bras, R., et al., Unsupervised Commonsense Question Answering with Self-talk, *arXiv:2004.05483*, 2020. DOI: https://doi.org/10.48550/arXiv.2004.05483

This paper was recommended for publication by RAS Academician D.A. Novikov, a member of the Editorial Board.

Received July 8, 2025, and revised November 17, 2025. Accepted December 16, 2025.

Author information

Kulinich, Aleksandr Alekseevich. Cand. Sci. (Eng.), Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia
✉ alexkul@rambler.ru
ORCID ID: https://orcid.org/0000-0002-4751-205X

Cite this paper

Kulinich, A.A., Application of Large Language Models in Decision Support Systems. Part I: Explanation Models and Large Language Models. *Control Sciences* **1,** 41–56 (2026).

Translated into English by *Alexander Yu. Mazurov*,
Cand. Sci. (Phys.–Math.),
Trapeznikov Institute of Control Sciences,
Russian Academy of Sciences, Moscow, Russia
✉ alexander.mazurov08@gmail.com