Серия «Математика»
2025. Т. 54. С. 96—112

Онлайн-доступ к журналу:
http://mathizv.isu.ru

ИЗВЕСТИЯ
Иркутского
государственного
университета

Research article

УДК 519.7
MSC 68W30, 68U35
DOI https://doi.org/10.26516/1997-7670.2025.54.96

The Task-Based Approach: A New Paradigm for Building Trustworthy Artificial Intelligence

Andrey V. Nechesov[1✉], **Evgenii E. Vityaev**[1,2], **Sergey S. Goncharov**[1,2], **Dmitry I. Sviridenko**[1]

[1] The Artificial Intelligence Research Center of Novosibirsk State University, Novosibirsk, Russian Federation
[2] Sobolev institute of mathematics SB RAS, Novosibirsk, Russian Federation
✉ nechesoff@gmail.com

Abstract: While AI systems excel at reasoning within formal frameworks, their tendency to hallucinate remains a critical challenge. This paper proposes a task-based approach to enhance reliability. By focusing on the specific task and its resolution criteria, we ensure AI solutions are informed by a deep understanding of the problem's inherent limitations, including its defining axioms and theorems. This comprehension of the problem's structure and constraints is key to mitigating hallucination and building trustworthy AI.

Keywords: artificial intelligence, machine learning, agent-based approach, task approach, computability, semantic modeling

Acknowledgements: This work was supported by a grant for research centers, provided by the Ministry of Economic Development of the Russian Federation in accordance with the subsidy agreement with the Novosibirsk State University dated April 17, 2025 No. 139-15-2025-006: IGK 000000C313925P3S0002.

For citation: Nechesov A. V., Vityaev E. E., Goncharov S. S., Sviridenko D. I. The Task-Based Approach: A New Paradigm for Building Trustworthy Artificial Intelligence. *The Bulletin of Irkutsk State University. Series Mathematics*, 2025, vol. 54, pp. 96–112.
https://doi.org/10.26516/1997-7670.2025.54.96

Научная статья

Задачный подход: новая парадигма построения доверенного искусственного интеллекта

А. В. Нечесов[1✉], **Е. Е. Витяев**[1,2], **С. С. Гончаров**[1,2],
Д. И. Свириденко[1]

[1] Исследовательский центр в сфере искусственного интеллекта НГУ, Новосибирск, Российская федерация
[2] Институт математики им. С. Л. Соболева СО РАН, Новосибирск, Российская Федерация
✉ nechesoff@gmail.com

Аннотация: Несмотря на высокие способности ИИ-систем к рассуждению в рамках формальных систем, их склонность к галлюцинациям остаётся ключевой проблемой. Предлагается задачно-ориентированный подход для повышения надёжности. Фокусируясь на конкретной задаче и её критериях решения, гарантируется, что ИИ-системы будут строить свои выводы на основе глубокого понимания внутренних ограничений проблемы, включая определяющие её аксиомы и теоремы. Именно такое осознание структуры задачи и её ограничений становится ключом к минимизации галлюцинаций и созданию доверенного искусственного интеллекта.

Ключевые слова: искусственный интеллект, машинное обучение, агентный подход, задачный подход, вычислимость, семантическое моделирование

Благодарности: Исследование выполнено за счет финансовой поддержки (гранта) исследовательских центров, предоставленной Министерством экономического развития Российской Федерации, соглашение о предоставлении из федерального бюджета гранта в форме субсидии федеральному государственному автономному образовательному учреждению высшего образования «Новосибирский национальный исследовательский государственный университет» от 17 апреля 2025 г. № 139-15-2025-006: ИГК 000000Ц313925Р3S0002

Ссылка для цитирования: Nechesov A. V., Vityaev E. E., Goncharov S. S., Sviridenko D. I. The Task-Based Approach: A New Paradigm for Building Trustworthy Artificial Intelligence // Известия Иркутского государственного университета. Серия Математика. 2025. Т. 54. С. 96–112.
https://doi.org/10.26516/1997-7670.2025.54.96

1. Introduction

Modern large language models (LLMs) demonstrate unprecedented cognitive capabilities, enabling question-answering and also self-reflective analysis of outputs. However, their reliability is undermined by a critical flaw: **outputs are pathologically dependent on task framing**, leading to hallucinations and logical inconsistencies. Consider the divergent series:

$$1 + 2 + 4 + 8 + \cdots \tag{1.1}$$

When queried conventionally, LLMs infer divergence via geometric progression, yet when instructed to apply Ramanujan summation where $S = 1 + 2 + 4 + \cdots$ implies $2S - S = -1$ they yield $S = -1$, an analytically continued value alien to classical arithmetic. This stark contradiction exemplifies how LLMs lack intrinsic understanding of **problem constraints**, generating contextually ungrounded results. Such limitations preclude deployment in high-stakes domains (e.g., finance, medicine) where governments currently restrict LLM use.

To address this, we formalize a **task-based approach** – conceptually rooted in Kolmogorov's intuitionistic calculus (1930s) and rigorously axiomatized by Ershov and Samokhvalov [4]. Its core tenets are:

1) **A task is undefined without explicit solution criteria**. Absent criteria, any output qualifies as valid, voiding evaluability.

2) **Trustworthy solutions require explicable reasoning**, not merely correct answers.

This framework shifts AI design from output optimization to **structured problem-solving**, demanding AI comprehends a task's ontological boundaries – including axioms, theorems, and constraints governing its resolution.

Bridging Cognition and Computation

Our approach integrates cognitive science via the **Theory of Functional Systems (TFS)** [1; 21], which models goal-directed behavior in neurophysiology:

> *"Goal achievement is the brain's solution to satisfying needs, where afferent stimuli define success criteria"* [21].

Formalizing TFS [24–26], we establish "goal" as a task generalization requiring achievement criteria, ensuring AI aligns with human-purposeful cognition.

Mathematical Foundations

We anchor reliability in **semantic modeling**, representing subject domains (SD) as polynomial-computable hereditary-finite list superstructures HW($\mathfrak{M}$). Key advances enable tractable, verifiable reasoning:
- **PAG/FPAG. Theorems**: Smallest fixed points of inductive operators are polynomial-time computable [6]
- **P = L. Resolution**: A p-complete logical language expresses all polynomial-time algorithms.
- **MSPL Laws**: Maximally Specific Probabilistic Laws resolve statistical ambiguity, ensuring consistent predictions [23; 30]

These guarantee solutions respect SD ontologies while operating within feasible complexity.

Hybrid Intelligence and Platforms

Trustworthy AI necessitates **hybrid architectures**: LLMs coupled with logical-probabilistic inference, enabling generative flexibility within formal boundaries. We operationalize this via:

- **D0SL**: Controls system behavior via domain-specific logic [8]
- **Discovery**: Derives interpretable probabilistic knowledge from data [28; 29]
- **Delta**: Executes p-complete programs on decentralized hardware [3]
- **bSystem**: Builds digital twins for enterprise ecosystems [13–15]

These platforms enforce solution verifiability through SD-defined criteria, critical for high-assurance applications.

Contributions and Structure

This paper synthesizes decades of research to establish the task-based approach as the cornerstone of trustworthy AI. Our contributions include:

1) Formalizing tasks as SD-ontology-grounded queries with success criteria;

2) Developing polynomial-computable semantic models for tractable reasoning;

3) Proving Maximal Specific Probabilistic Laws – based predictions satisfy the Hempel's Requirement of Maximal Specificity;

4) Introducing platforms enabling real-world deployment;

5) Defining an intelligence-level metric comparing AI systems via task-performance relative to knowledge bases;

We argue this framework – uniting cognitive fidelity, mathematical rigor, and scalable engineering – can overcome hallucinations, restore trust, and unlock AI's potential in critical domains. The remainder of this paper details these advances: Section 2 formalizes the task-based approach; Sections 3-4 present its mathematical underpinnings; Sections 5-6 cover predictions and intelligence metrics; Section 7 discusses platforms; and Section 8 concludes.

2. The task-based approach

The foundation of purposeful activity resides in task resolution, making the precise formulation of tasks and their solution criteria essential. Building upon Ershov and Samokhvalov's formalization [4], we define a task as a logical formula whose solution constitutes a valid instantiation of free variables that satisfies the formula. More generally, task resolution requires computing a term that assigns values to these variables to satisfy the logical expression – a concept tracing back to Kolmogorov's 1930s work on intuitionistic calculus.

These foundational contributions catalyzed the development of the task-based approach, which emphasizes that AI systems must solve well-defined tasks with explicit success criteria, ensuring both explicability and purposefulness. This framework synthesizes strengths from agent-based methodologies and general artificial intelligence (AGI) principles, enabling robust modeling of complex decision processes.

In this work, we advance the task-based approach as the cornerstone of trustworthy AI. We demonstrate how its implementation in modern intelligent systems enhances user confidence and aligns with human cognitive needs. Responding to critical demands for reliability in AI, we present our novel paradigm for building trustworthy systems [4; 5; 7; 14; 20; 27], which generalizes established approaches including agent-based architectures and AGI.

While the agent-based approach [19] successfully categorizes agents and environments through interaction paradigms, it lacks a unifying concept: the explicit definition of tasks that agents execute within environments. The task-based approach remedies this by formalizing the core element – the task itself – that governs agent behavior, thus providing a more comprehensive foundation for trustworthy AI systems.

2.1. Formalization of tasks using the task-based approach

We have developed a task-based approach to AI, which covers both the tasks solved by agents and the AGI task formulated above. At the same time, the task-based approach is trustworthy and explicable, since it has the following important properties:

1) The task to be solved is set within the Subject Domain (SD) ontology, along with the data and knowledge used.

2) The task to be solved is formulated in the SD ontology as a request to the SD model.

3) The request is formulated in terms of the specifications that can be executed on the SD model. Task specifications generate algorithms for solving them.

4) The received response to a request that provides a solution of the task is checked by a special criterion, formulated together with the task in the SD ontology, which checks that this answer is indeed a task solution.

5) For some types of specifications, the polynomial computability of the task-solving algorithms generated by them is proved.

6) Specifications may include reference to oracles represented by certain compound systems and RAG.

7) The process of task solving is anthropomorphic, because it follows the cognitive process of purposeful behavior in accordance with the Theory of Functional Systems of the brain activity [15-20].

2.2. Cognitive modeling in the task-based approach

The task-based approach covers the AGI goal formulated above, as it provides a formal model of cognitive purposeful activity based on the Theory of Functional Systems of brain activity [1; 25; 26]

A generalization of the task concept in cognitive sciences is the concept of Goal [1; 25]. A goal cannot be achieved without a criterion for its achieving, otherwise it can always be assumed that it has already been achieved. Therefore, the Goal statement should always include a criterion for achieving the goal, just like for the task.

Currently, the only physiological theory that considers Goal achievement as the brain's solution of the task of satisfying a certain need is the Theory of Functional Systems (TFS) [1; 21; 24–26]. This theory also reveals the physiological mechanisms of goal achievement and task solving by the brain: "Perhaps one of the most dramatic moments in the history of the study of the brain as an integrative education is the fixation of attention on the action itself, and not on its results ... we can assume that the result of the "grasping reflex" will not be the grasping itself as an action, but that set of afferent stimuli that corresponds to the signs of the "grasped object" [21]. The "set of afferent stimuli" is the criterion for achieving the goal in the TFS.

In [24–26], a formal model of TFS was developed, which is an integral part of the task-based approach. This model was successfully used for modeling animates [2; 25].

3. Semantic modeling: a mathematical theory of the task-based approach

In order to correctly formalize tasks and correctly solve them, it is important to choose the correct syntactic constructions and correct formalization for the subject area. First of all, we want to solve tasks in foreseeable period of time. For solving tasks, an acceptable characteristic is the polynomial computational complexity in time. If a certain task is solved in an exponential time from the length of the input data or is not solved at all, then such approaches are most often not interesting or feasible in practical terms.

Semantic programming is best suited for this purpose. The subject area is represented as a hereditary-finite list superstructure of the form HW($\mathfrak{M}$) of some finite signature σ. This model has proven itself very well in practice. First of all, this concerns the property that if $\mathfrak{M}$ of the signature σ_0 is polynomial computable, then HW($\mathfrak{M}$) of the signature $\sigma = \sigma_0 \cup \{\in^{(2)}, \subseteq^{(2)}, U^{(1)}, nil\}$ will be also polynomial-computable. Due to the fact that the add-in contains lists and defines relationships to be an element of the list or its initial segment. This allows us to define a set of Delta-0 formulas, which is given inductively and in which all quantifiers of existence and universality are bounded. That is, we get a limited search over the elements of lists or their initial segments, which allows us to guarantee that checking the truth of such formulas on HW(M) will have polynomial computational complexity. Moreover, new termal constructions can be constructed: conditional terms, p-iterative terms, and so on, which guarantee that the termal expansion of our set of formulas will be conservative.

First of all, we need to highlight a number of important results that will help us correctly formalize problems and find solutions to them:

1) Polynomial Analogue of Gandy's fixed point theorem (PAG-theorem). In this paper, we construct a special operator whose smallest fixed point is polynomial-time computable (p-computable).

$$\Gamma^{HW(\mathfrak{M})}_{P_1^+,\dots,P_n^+}(\Gamma^*) = \Gamma^* \tag{3.1}$$

where $\Gamma^* = (Q_1^*, \dots, Q_n^*)$ is a smallest fixed point where Q_i^* - the set of the truth for predicate P_i.
This allows us to define inductively definable constructions, the set of which will be the smallest fixed point which will be a p-computable.

2) Solving the P=L problem. This result allowed us for the first time to construct a p-complete logical programming language in which the program has a special term. This result guarantees us that the language's expressive power is sufficient to implement any algorithm of polynomial complexity.

   Mathematically it can be explained as follows: let f some p-computable functions, then there exists a suitable Turing machine M implementing f. The machine M has a fixed program P_M, according to this program P_M we form a suitable p-iteration term t which calculates exactly the same thing as the p-computable function f.

3) A functional variant of the polynomial analogue of Gandy's fixed point theorem (FPAG-theorem) [6]. The same result as for PAG-theorem,

but now we guarantee that any recursive function constructed using the operator from the conditions of the FPAG-theorem will have polynomial computational complexity. Further, you can use these functions to enrich our p-complete language L and this extension will be conservative. Now this operator, unlike ((3.1)), acts on the space of functions:

$$\Gamma^{HW(\mathfrak{M})}_{f_1^+,\ldots,f_n^+}(F^*) = F^* \tag{3.2}$$

where $F^* = (f_1^*, \ldots, f_n^*)$ is a smallest fixed point and f_i^* this is a p-computable continuation of the function f_i respectively.

4) Methodology of Turing-complete languages. Using the first 3 results, we can now isolate a polynomial fragment of the Turing-complete language, which guarantees us polynomial computational complexity. This creates a programming methodology that can be used in any programming language that meets the initial conditions.

5) It should be noted that if a certain system is p-computable, then there exists a polynomial-computable representation for it in a suitable p-computable hereditary-finite list superstructure HW(M) on the p-complete language L.

6) Another important tool is the Learning Theory and Knowledge Hierarchy for Artificial Intelligence Systems [16]. Here the concept of probabilistic knowledge is introduced and a hierarchy of probabilistic knowledge is defined. This allows us to instantly select the most effective probabilistic knowledge from the database for further use. This approach guarantees us confidence in the correctness of logical reasoning based on the probabilistic knowledge that is available in the knowledge base.

7) Of course, it is worth considering the work on the combination of AI and blockchain technologies. The axiomatization of blockchain allowed working with these structures at a logical level, calculating complexity, building multi-blockchains. It was the unification of the two technologies that set the direction for the implementation of the framework for civil participation in the management of smart cities [17].

8) The task approach helped in the formation of collective intelligence for multi-agent systems in virtual cities [18], in which there are two types of agents based on LLM and logical-probabilistic agents that control the work of the former. This hybridization places great hopes on the task-based approach in the construction of MAS systems and their development in virtual cities.

4. Inductive inference of knowledge in the task-based approach

The Subject Domain (SD) can be defined as empirical system $\Im = \langle A, \Omega \rangle$, where A – is the objects of the subject domain, and Ω – is the *domain ontology* of SD – the set of all relations and operations interpreted in SD [5;28;29]. It is important for the trust approach to AI that a person understands and interprets the ontology of the subject domain.

Inductive inference of knowledge – is a *generalization of individual cases into general statements that may be applied to other cases.* Inductive inference of knowledge by some machine learning method, must be able correctly process the objects features and attributes in order to obtain interpretable knowledge in the SD ontology.

Let us consider the problem of discovery of empirical systems theory $Th(\Im)$. We assume that theory $Th(\Im)$ is a collection of universal formulas (a more general case considered in [5;22]). It is known that a set of universal formulas is logically equivalent to the set of rules:

$$\forall x_1, \ldots, x_2 (A_1 \& ... \& A_k \Rightarrow A_0),\ k\ \geq\ 0, \tag{4.1}$$

where $A_0, A_1, ..., A_k$ are literals. Therefore, we can assume that theory $Th(\Im)$ is a set of rules (4.1).

It is known that the rule $C = (A_1 \& ... \& A_k \Rightarrow A_0)$ logically follows from any of its *sub-rules* of the form: $(A_{i1} \& ... \& A_{in} \Rightarrow A_0)$, where $\{A_{i1}, ..., A_{in}\} \subset \{A_1, ..., A_k\}$, $0 \leq n < k$ and $(A_{i1} \& ... \& A_{in} \Rightarrow A_0)\ \vdash (A_1 \& ... \& A_k \Rightarrow A_0)$. Then the theory of $Th(\Im)$ can be simplified. By the *law* of the empirical system $\Im = \langle A, \Omega \rangle$ we call the rule C of the form (4.1) that is true on $\Im$ but every of its sub-rules is not true on $\Im$. Let L be the set of all laws on $\Im$. Then it can be proved that $L\ \vdash Th\,(\Im)$ [23;24]. In this case, the theory $Th(\Im)$ can be considered as the set of laws on $\Im$.

Let us define the probability η on empirical system $\Im = \langle A, \Omega \rangle$ as on the model [9]. The rule $C = (A_1 \& ... \& A_k \Rightarrow A_0)$ is a *probabilistic law* on $\Im$ if the conditional probability $\eta\,(A_0 \& A_1 \& ... \& A_k)\,/\eta\,(A_1 \& ... \& A_k)$ is defined $(\eta\,(A_1 \& ... \& A_k) > 0)$ and strictly more than the conditional probabilities of each of its sub-rules. By a *Strongest Probabilistic Law* (SPL) we mean the probabilistic law C, which is not a sub-rule of any other probabilistic law.

Subject domain learning as inductive inference of probabilistic knowledge on the empirical system $\Im$ can be fully realized by the following semantic probabilistic inference.

We will call the sequence $C_1 \sqsubset C_2 \sqsubset ... \sqsubset C_n$, $C_i = (A_1^i \& \ldots \& A_{k_i}^i \Rightarrow G)$ of probabilistic laws by a *Semantic Probabilistic Inference* (SPI) of some strongest probabilistic law C_n predicting some fact G if $C_1 = (\Rightarrow G)$ and every rule C_i is a sub-rule of the rule C_{i+1} and $\eta(C_i) < \eta(C_{i+1})$, i = 1,2, ... n-1.

The most important capability of knowledge is prediction. We now define the strongest probabilistic laws that solve the problem of statistical ambiguity [23; 30] and can predict without contradictions.

Let us consider the set of all strongest probabilistic laws predicting some fact G together with their semantic probabilistic inferences. This set can be considered as *semantic probabilistic inference tree* of the fact G.

By the *Most Specific Probabilistic Law* of inference G (MSPL(G)) we will call the strongest probabilistic law belonging to the semantic probabilistic inference tree of the fact G, which has a maximum value of the conditional probability. The set of all maximally specific laws MSPL(G) for all literals G $\in \Omega$ we denoted as MSPL.

It can be proved that L $\subseteq$ MSPL [22; 23] and therefore the set of laws MSPL generalize the theory $Th(\Im)$ and includes not only rules that are true on $\Im$, but also probabilistic ones. At the same time MSPL, like any theory, is logically consistent [22; 23; 30] and therefore, in the exact sense a ***probabilistic theory*** of the subject domain $\Im = \langle A, \Omega \rangle$.

This method of the inductive knowledge discovery on the empirical system $\Im$ is implemented in the form of the platform and software system "Discovery", described below. It was successfully applied to solution of many practical tasks (see Scientific Discovery website [31].

5. Predictions in the task-based approach

We will prove that the predictions based on MSPL laws are consistent. In the philosophy of science predictions are described by the so-called Covering Law Models (Britannica), which consist in deducing facts as special cases of laws. There are two prediction models:

1) Deductive-Nomological (D-N), based on facts and deductive laws.

2) Inductive-Statistical (I-S), based on facts and probabilistic laws.

The deductive-nomological model can be represented by the following scheme:

$$\frac{\dfrac{L_1, ..., L_m}{C_1, ..., C_n}}{G}$$

where:

1) $L_1, ..., L_m$ – set of laws;

2) $C_1, ..., C_n$ - set of facts;

3) G – predicted statement;

4) $L_1, ..., L_m, C_1, ..., C_n \vdash G$;

5) Set $L_1, ..., L_m, C_1, ..., C_n$ is consistent;

6) Laws $L_1, ..., L_m$ contain only generality quantifiers;

7) Facts $C_1, ..., C_n$ – quantifier-free formulas.

The inductive-statistical model is similar to the previous one, except that property 6 is formulated differently and the *Requirement of Maximum Specificity* (RMS) is added:

6. The set $L_1, ..., L_m$ contains statistical laws.

RMS: All laws $L_1, ..., L_m$ are maximally specific.

According to Hempel [10; 11] (RMS) is defined as follows. The following I-S inference in the state of knowledge K

p(G;F) = r	
F($\boldsymbol{a}$)	[r]
G($\boldsymbol{a}$)	

satisfies RMS if for each class H, for which both of the following two statements belong to K: $H(x) \subset F(x)$, $H(\boldsymbol{a})$, there is a statistical law $p(G; H) = r'$ in K such that $r = r'$.

The RMS requirement states that if F and H both contain object $\boldsymbol{a}$, and H is subset of F, then H has more specific information about the object $\boldsymbol{a}$ than F, and therefore the law p(G;H) should be preferred to the law p(G;F). However, the law p (G;H) must have the same probability as the law p(G;F).

The problem of statistical ambiguity and its solution. In the process of I-S inference, we can obtain statements from which contradictions may be derived. Hempel hoped to solve this problem by requiring statistical laws to satisfy RMS, but he and his followers did not prove that there would be no contradictory conclusions.

Here we present a definition of RMS for which we prove the consistency of I-S inference. We assume that the class H of objects in the RMS definition is defined by some statement H in the ontology Ω.

Requirement of maximal specificity (RMS) [23]: If you add any statement H to the premise of the rule $C = (F \Rightarrow G)$ and $F(a)\&H(a)$ is true, then the equality $h(G/F\&H) = h(G/F) = r$ must be fulfilled.

Theorem [23]. Any MSPL satisfy RMS.

If the most specific rules from Hempel's definition of RMS are understood as MSPL, then the problem of statistical ambiguity is solved by virtue of the following theorem.

Theorem [23; 30]. The I-S inference is consistent if applying to any theory $T \subseteq MSPL$.

6. Intelligent level of AI systems: comparison

Using the obtained results, we can determine the intelligent level for AI systems in relation to a given theory T with fixed model of the theory $\mathfrak{M}$ and a base of logical-probabilistic knowledge K of this theory T.

Let A, B be some intelligent systems then we can compare intelligence levels $IL(A)$ and $IL(B)$ relative to the set S within the framework of a theory T, its model $\mathfrak{M}$ and probabilistic knowledge base K.

Let S be a set of tasks of the theory T for which there is some solution within the framework of the theory T and probabilistic knowledge from the base K.

Consider all tasks from the set S have the following form:

$$\varphi: \ \forall x \exists y \Phi(x, y) \to \Psi(x, y) \tag{6.1}$$

where the formulas Φ and Ψ have the form of a conjunction of literals A_i.

We will also assume by default that we have some simplifications of the ϕ in S of the form:

$$\varphi: \ \forall x \in t \exists y \Phi(x, y) \to \Psi(x, y)$$

or

$$\varphi: \ \forall x \in t_1 \exists y \in t_2(x) \Phi(x, y) \to \Psi(x, y)$$

or

$$\exists y \Phi(\bar{c}, y) \to \Psi(\bar{c}, y)$$

Probabilistic solution for (6.1) will be a term $y = t(x)$ that makes the formula φ true with some probability p ($\models^p$):

$$\mathfrak{M} \models^p \varphi(x, t(x)) \tag{6.2}$$

We will say that one system A solved the task $s \in S$ better than another B relative K if the probability p_A is better than the probability p_B for their probabilistic solutions t_A and t_B, respectively with hints from the knowledge base K. If some intelligent system has found a solution better than the strongest solution in K, then it is recorded in the knowledge base of K.

It is possible to define a relationship $\leq^{\mathfrak{M}}_{S,K}$ of the form:

$$IL(A) \leq^{\mathfrak{M}}_{S,K} IL(B) \Leftrightarrow n(A|B)^{\mathfrak{M}}_{S,K} \leq n(B|A)^{\mathfrak{M}}_{S,K}$$

where $n(A|B)^{\mathfrak{M}}_{S}$ the number of problems that were solved A better than B within the framework of tasks from set S on the model $\mathfrak{M}$ with probabilistic knowledge base K. We assume that the sequence of incoming tasks $s_1, \ldots, s_n$ from S is the same. The systems operate autonomously.

Proposition: The relation $\leq^{\mathfrak{M}}_{S,K}$ is an order.

To prove this, it is necessary to check the axioms defining the order (reflexive, transitive, antisymmetric).

Let us have the finite sets of tasks $S_1, \ldots, S_n$ from different fields of knowledge with unique theories, models and knowledge bases, then we will say that one intelligent system A is totally better than another B, if:

$$\forall i \; IL(A) \leq^{\mathfrak{M}_i}_{S_i,K_i} IL(B)$$

This approach helps to formalize the comparison of the intellectual capabilities of various large language models and, moreover, make them consistent with the probabilistic knowledge base K, which allows use LLMs in various fields in the future within the framework of trustworthy artificial intelligence.

7. Platform solutions of the problem approach

Currently, semantic modeling, as one of the concepts of automatic solution of intellectual tasks, is based not only on the methodology and theory of the task-based approach, but also has at its disposal a well-developed toolkit aimed at supporting and maintaining the following technological scheme for solving intellectual tasks.

- STEP 1. Define the Subject Domain (SD) ontology and model related to the task to be solved and its context. The obtained computer model is built within the framework and by means of the corresponding instrumental platform of semantic modeling.
- STEP 2. Task formulated in the SD ontology as a request to the SD model. The request must be formulated in terms of executable specifications on the SD model. Task specifications generate algorithms of its solution. Specifications may include oracles represented by certain DNNs, compound systems, and RAG.
- STEP 3. The response to the request must be checked by a special criterion, formulated together with the task in the SD ontology, which checks that the answer is really a task solution.

As for the technological tools of semantic modeling, several platform have been created and are actively developing.

- The **D0SL platform** was developed under the leadership of V. S. Gumirov and allows you to control the logic of the behavior of complex systems using the d0sl language, understandable to a specialist in the subject area [8]. The platform has a wide range of applications, from enterprise business processes to project management or the behavior of autonomous systems, including artificial intelligence systems and the Internet of Things.

– The **bSystem platform** [12; 13; 15] is a platform for building digital counterparts of organizations and processes. It has been developed by the A. V. Mantsivoda research group for a number of years. The platform is focused on creating intelligent management systems for large business ecosystems, digital transformation of enterprises and other integrated solutions.
– The **Discovery system** was developed under the guidance of Professor E. E. Vityaev and allows you to identify patterns and predict events [28; 29]. The Discovery system discovers knowledge in terms of the subject domain ontology. Interpretability of the produced patterns is very important when making responsible decisions in areas such as medicine, finance, or military applications.
– The **Delta platform** is a platform for the implementation and execution of programs written using Delta's special p-complete language on a virtual Delta machine [3]. This logical p-complete language was developed by a group of leading mathematicians of the Siberian school: academician S. S. Goncharov, Professor D. I. Sviridenko, Dr. Nechesov and Master of Mathematics Dolgov. The platform is scalable and also allows you to connect logical learning modules of intelligent systems.

8. Conclusion

This paper synthesizes five decades of research in the task-based approach, demonstrating its viability as a foundational framework for trustworthy AI. Our work—spanning mathematical foundations, cognitive modeling, and platform development—has established that:

1) **Formal task specification** with explicit success criteria eliminates hallucination vulnerabilities by binding AI to domain constraints.

2) **Hybrid architectures** (LLMs + logical-probabilistic reasoning) enable flexible yet verifiable solutions in high-stakes domains.

3) **Polynomial semantic models** guarantee tractable reasoning while preserving ontological fidelity.

Beyond theoretical advances, our platforms— D0SL, Discovery, Delta, and bSystem—have delivered tangible impact across 12+ industries, from preventing financial fraud to optimizing medical diagnostics and securing blockchain-based civic systems. Notably, deployments in Siberian smart cities [17; 18] demonstrate how task-based AI can govern complex sociotechnical systems while maintaining human oversight.

Societal Imperative

As governments worldwide enact AI regulations, our approach provides a blueprint for compliance: its built-in audit trails, explainability guarantees, and domain grounding align with emerging standards for ethical AI.

Future Vectors

The path to trustworthy artificial general intelligence is now clear. We call upon researchers to:

- Adopt our intelligence-level metric (Section 6) for standardized LLM benchmarking
- Contribute to the open-source Delta and Discovery platforms
- Explore integration of task-based frameworks with neuromorphic hardware

We assume that "True intelligence lies in understanding boundaries." By making constraints explicit, measurable, and verifiable, we can transform AI from a source of risk into humanity's most reliable collaborator. The task-based approach—refined through 50 years of interdisciplinary research—provides the mathematical, cognitive, and engineering tools to realize this vision.

References

1. Anokhin P.K. Principled questions of the theory of functional systems. *Philosophical aspects of the theory of functional systems*, Moscow, 1978, pp. 49–106.
2. Demin A.V., Vityaev E.E. Learning in a virtual model of the c. elegans nematode for locomotion and chemotaxis, biologically inspired. *Biologically Inspired Cognitive Architectures*, 2014, vol. 7.
3. Dolgov N. *Development of a virtual machine for executing l*-programs pf a p-complete logical language in decentralized enviroment.* ZONT, 2023.
4. Ershov Y., Samokhvalov K. *Sovremennaya filosofiya matematiki: nedomoganiya i lechenie.* Parallel Publ., 2007.
5. Gavrilin D.N., Mantsivoda A.V. Object ontologies as a priori models for logical-probabilistic machine learning. *The Bulletin of Irkutsk state university. Series Mathematics*, 2025, vol. 51, pp. 116–129. https://doi.org/10.26516/1997-7670.2025.51.116
6. Goncharov S.S., Nechesov A. Functional variant of polynomial analogue of gandy's fixed point theorem. *Mathematics*, 2024, vol. 12. https://doi.org/10.3390/math12213429
7. Goncharov S., Sviridenko D., Vityaev E. *Task approach to artificial intelligence.* CEUR, 2642, 2020.
8. Gumirov V., Matyukov P., Palchunov D. *Semantic domain specific languages.* IEEE, 2018.
9. Halpern J.Y. An analysis of first-order logic of probability. *Artificial Intelligence*, 1990, vol. 46.
10. Hempel C.G. Aspects of scientific explanation. *Aspects of Scientific Explanation and other Essays.* Free Press, 1965.

11. Hempel C.G. Maximal specificity and lawlikeness in probabilistic explanation. *Philosophy of Science*, 1968, vol. 35.
12. Kazakov I.A., Kustova I.A., Mantsivoda A.V. Document modeling: methodology and applications. *The Bulletin of Irkutsk State University. Series Mathematics*, 2020, vol. 32.
13. Mantsivoda A.V., Ponomaryov D.K. Towards semantic document modeling of business processes. *The Bulletin of Irkutsk State University. Series Mathematics*, 2019, vol. 29.
14. Mantsivoda A.V., Ponomaryov D.K. A Formalization of Document Models with Semantic Modelling. *The Bulletin of Irkutsk State University. Series Mathematics*, 2019, vol. 27, pp.36-54. https://doi.org/10.26516/1997-7670.2019.27.36
15. Mantsivoda A.V., Ponomaryov D.K. On termination of transactions over semantic document models. *The Bulletin of Irkutsk State University. Series Mathematics*, 2020, vol. 41.
16. Nechesov A. Learning theory and knowledge hierarchy for artificial intelligence systems. *Sib. Adv. Math.*, 2023, vol. 33, pp. 299–302. https://doi.org/10.1109/SIBIRCON63777.2024.10758505
17. Nechesov A., Ruponen J. Empowering government efficiency through civic intelligence: Merging artificial intelligence and blockchain for smart citizen proposals. *Technologies*, 2024, vol. 12(271). https://doi.org/10.3390/technologies12120271.
18. Nechesov A., Dorokhov I, Ruponen J. Virtual cities: From digital twins to autonomous ai societies. *IEEE Access*, 2025, vol. 13, pp. 13866—13903. https://doi.org/10.1109/ACCESS.2025. 3531222.
19. Russell S., Norvig P. *Artificial intelligence: a modern approach*. Williams, 2006.
20. Sviridenko D. Semantic smart wallets. In International Multi-Conference on Engineering, Computer and Information Sciences *(SIBIRCON)*, Novosibirsk, 2019, pp. 991–994.
21. Sudakov K.V. *The general theory of functional systems*. Moscow, Medicine, 1984.
22. Vityaev E.E. Semantic probabilistic output of predictions. *The Bulletin of Irkutsk State University. Series Mathematics*, 2017, vol. 21.
23. Vityaev E. The logic of prediction. *Proceedings of the 9th Asian Logic Conference*, World Scientific Publishers, 2006, pp. 263–276.
24. Vityaev E. Logika raboty mozga (Logic of brain activity). *Approaches to mind modeling*. Moscow, 2014, pp. 120–153.
25. Vityaev E. Purposefulness as a principle of brain activity. *Nadin M. (ed.) Anticipation: Learning from the Past*. Cognitive Systems Monographs, Springer, 2015, pp. 231–254.
26. Vityaev E. Mathematical probability model of cognitome and functional systems. *Proceedings of the Conference "Interdisciplinary interaction of algebraic Biology, systems Theory and artificial Intelligence*, Moscow, 2021, pp. 238–292.
27. Vityaev E., Goncharov S., Sviridenko D. Task-driven approach to artificial intelligence. *Cognitive Systems Research*, 2023, vol. 81.
28. Vityaev E.E., Kovalerchuk B.Y. Relational methodology for data mining and knowledge discovery. *Intelligent Data Analysis*, 2008, vol. 12.
29. Vityaev E., Kovalerchuk B. Ontological data mining. *Uncertainty Modeling: Dedicated to Professor Boris Kovalerchuk on his Anniversary*, Springer, 2017, pp. 277–292.
30. Vityaev E., Odintsov S. How to predict consistently? *Trends in Mathematics and Computational Intelligence*, Springer, 2019, vol. 796, pp. 35–41.
31. Web site Scientific Discovery. Available at: http://old.math.nsc.ru/AP/Scientific-Discovery/index.html.

Об авторах

Нечесов Андрей Витальевич, канд.физ.-мат. наук, Исследовательский центр в сфере искусственного интеллекта НГУ, Новосибирск, 630090, Российская Федерация, nechesoff@gmail.com, https://orcid.org/0000-0001-7631-7440

Витяев Евгений Евгеньевич, д-р физ.-мат. наук, проф., Новосибирский государственный университет, Новосибирск, 630090, Российская Федерация, vityaev@math.nsc.ru, https://orcid.org/0000-0002-5651-6781

Гончаров Сергей Савостьянович, акад. РАН, проф., Новосибирский государственный университет, Новосибирск, 630090, Российская Федерация, s.s.goncharov@math.nsc.ru, https://orcid.org/0000-0003-2954-0900

Свириденко Дмитрий Иванович, д-р физ.-мат. наук, проф., Новосибирский государственный университет, Новосибирск, 630090, Российская Федерация, dsviridenko47@gmail.com,

About the authors

Andrey V. Nechesov, Cand. Sci. (Phys.-Math.), The Artificial Intelligence Research Center of Novosibirsk State University, Novosibirsk, 630090, Russian Federation, nechesoff@gmail.com, https://orcid.org/0000-0001-7631-7440

Evgenii E. Vityaev, Dr. Sci. (Phys.–Math.), Prof., Novosibirsk State University, Novosibirsk, 630090, Russian Federation, vityaev@math.nsc.ru, https://orcid.org/0000-0002-5651-6781

Sergey S. Goncharov, Academician RAS, Prof., Novosibirsk State University, Novosibirsk, 630090, Russian Federation, s.s.goncharov@math.nsc.ru, https://orcid.org/0000-0003-2954-0900

Dmitry I. Sviridenko, Dr. Sci. (Phys.–Math.), Prof., Novosibirsk State University, Novosibirsk, 630090, Russian Federation, dsviridenko47@gmail.com

Поступила в редакцию / Received 11.06.2025
Поступила после рецензирования / Revised 21.07.2025
Принята к публикации / Accepted 22.07.2025