



## Повышение достоверности ответов диалоговой модели с использованием формальных грамматик и теории категорий

© 2026, В.В. Антонов<sup>1</sup>, Л.Е. Родионова<sup>1</sup>✉, Е.В. Пальчевский<sup>2</sup>, В.А.Суворова<sup>1</sup>

<sup>1</sup>Уфимский университет науки и технологий (УУНУТ), Уфа, Россия

<sup>2</sup>МИРЭА - Российский технологический университет (РТУ МИРЭА), Москва, Россия

### Аннотация

Работа посвящена повышению достоверности ответов диалоговой модели для задач, требующих высокой точности. Предложен гибридный подход, сочетающий дообучение модели на специализированных данных с интеграцией формальных грамматик Хомского в процесс обработки запроса. Предложена архитектура, объединяющая формальный синтаксический контроль с генеративными возможностями дообученной модели ГигаЧат 2.0. В процесс обработки запроса включён модуль, реализующий предварительный анализ и фильтрацию запросов на основе контекстно-свободных грамматик Хомского. Для согласования работы генеративной модели с выявленной синтаксической структурой предложен двухуровневый механизм контроля. На входе синтаксически-обогащённый запрос направляет процесс на структурные паттерны запроса. На выходе модуль проверки осуществляет выбор и коррекцию ответа на основе формальных критериев. Для повышения объяснимости и формального анализа потоков данных использована теория категорий, где запросы, промежуточные представления и ответы моделируются как морфизмы и объекты категории. Для улучшения логической связности и учёта контекста предложено геометрическое представление семантики в пространстве Минковского, метрика которого позволяет разделять инвариантное значение слова и его контекстуально-временную динамику. Экспериментальная оценка на специализированном наборе из 500 фактологических и структурированных запросов показала, что предложенный метод снижает уровень генерации вымышленных утверждений на 22% по сравнению с базовой моделью ГигаЧат 2.0. Показано, что совместное использование формальных грамматик, категориального формализма и геометрической модели семантики позволяет повысить достоверность ответов в предметно-ориентированных диалогах.

**Ключевые слова:** достоверность ответов, диалоговая модель, трансформер, формальные грамматики Хомского, теория категорий, пространство Минковского, семантическое пространство.

**Цитирование:** Антонов В.В., Родионова Л.Е., Пальчевский Е.В., Суворова В.А. Повышение достоверности ответов диалоговой модели с использованием формальных грамматик и теории категорий. *Онтология проектирования*. 2026. Т.16, №1(59). С.125-138. DOI: 10.18287/2223-9537-2026-16-1-125-138.

**Финансирование:** исследование выполнено при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках государственного задания высшим учебным заведениям № FRRR-2026-0006.

**Вклад авторов:** Антонов В.В. – разработка структуры статьи и подбор источников; Родионова Л.Е. – подготовка примеров; Пальчевский Е.В. – разработка диалоговой модели; Суворова В.А. – верификация результатов.

**Конфликт интересов:** авторы заявляют об отсутствии конфликта интересов.

### Введение

Современный этап развития науки о данных характеризуется прогрессом в области обработки естественного языка (ЕЯ), ключевым элементом которого стали большие языковые модели (БЯМ) на основе трансформеров (<https://giga.chat/>, <https://deepseek.com.ru/>, <https://openai.com/> и др.) [1]. Российская модель ГигаЧат 2.0 (<https://giga.chat/>) оптимизирова-

на для работы с русскоязычным контентом и предоставляет возможности для её дообучения, что обусловило её выбор в качестве базовой.

Основой развития трансформеров стало применение механизмов внимания, позволивших учитывать зависимости между словами и конструкциями текста. Благодаря этому удалось обеспечить «понимание» контекста и повысить точность моделей при выполнении сложных задач обработки ЕЯ. Примером является *Funnel-Transformer* [2], который показал превосходство над другими моделями в ряде задач, требующих анализа семантики текста. Особенностью данной модели является способность вначале сформировать обобщённое высокоуровневое представление контекста, а затем, при детальном анализе, извлекать необходимые подробности из первичного представления.

Распространение получили подходы, основанные на предобучении БЯМ, такие как *BERT*, *GPT* и др. Модели обучаются на больших объёмах неразмеченных данных, усваивая общие закономерности языка, и затем применяются для решения специфических задач. Это позволяет многократно использовать знания, заложенные в предобученных моделях, для адаптации их к прикладным задачам. Недостатком современных БЯМ является склонность к «галлюцинациям» – генерации правдоподобной, но фактически неверной (вымышленной) информации [1, 3]. В образовательных, экспертных и прикладных сценариях цена ошибки высока. Известные подходы к устранению проблемы требуют огромных затрат на разметку и не гарантируют достижения цели [4].

Методы повышения достоверности БЯМ можно разделить на классы с разным балансом между точностью, контролируемостью и вычислительной эффективностью.

- Тонкая настройка с использованием обучения с подкреплением на основе человеческих оценок (*RLHF*) – стандартный подход для моделей типа ГигаЧат [2, 4, 5].
- Генерация, дополненная поиском (*RAG*) – подход, при котором модели дополняют процесс генерации текста поиском информации во внешних базах знаний [5].
- Нейро-символьные гибриды - интеграция нейросетей с формальными системами [6, 7].
- Геометрические модели семантики - использование неевклидовых пространств (гиперболических, сферических) для представления иерархий и аналогий [8].
- Применение теории категорий в машинном обучении - категориальный подход используется для формализации архитектур, например, с применением математического аппарата к сложным структурам в [8].

Для диалоговых систем, основанных на БЯМ, необходимо обеспечить предсказуемость, контролируемость и достоверность генеративных компонент в условиях открытых предметных областей. Подходы к проектированию систем обработки ЕЯ, основанные на обучении на данных, не обеспечивают необходимый уровень формальной гарантии корректности. Существует потребность в методологиях проектирования, сочетающих гибкость нейросетевых моделей с детерминизмом формальных методов.

Настоящая работа направлена на разработку гибридного метода, который дополняет возможности генеративной трансформер-архитектуры (ГигаЧат 2.0) формальными механизмами контроля. В качестве таких механизмов предлагается использовать:

- формальные грамматики Хомского для синтаксического анализа и фильтрации входных запросов и выходных ответов;
- аппарат теории категорий для формализации процесса обработки данных и анализа потоков информации;
- геометрическую модель семантики на основе пространства Минковского для улучшения учёта контекста и логической связности.

## 1 Теоретические основы архитектуры трансформеров

Трансформеры представляют собой архитектурный паттерн для нейронных сетей, основанный на механизме самовнимания [9].

## 1.1 Общие положения

Базовый механизм внимания представляет собой следующее: для входной последовательности в виде матрицы векторные представления (слов, предложений, изображений и др.)  $X$  вычисляются три матрицы: запросов  $Q = X * W_Q$ , ключей  $K = X * W_K$  и значений  $V = X * W_V$ , где  $W_Q, W_K$  и  $W_V$  – обучаемые матрицы параметров (весов), выполняющие роль линейных проекций. Веса внимания рассчитываются по формуле (1) скалярного произведения с последующим применением функции *Softmax* [9]:

$$Attention(Q, K, V) = V \cdot \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right), \quad (1)$$

где  $d_k$  – размерность ключей,  $W_Q, W_K$  и  $W_V$  – обучаемые матрицы параметров (весов), выполняющие роль линейных проекций; масштабирующий множитель  $\frac{1}{\sqrt{d_k}}$  стабилизирует градиенты.



Рисунок 1 – Схема архитектуры гибридной системы с внешним контролем

Архитектура трансформера включает энкодер и декодер. Для задач генерации текста, включая диалоговые системы, наибольшее распространение получили авторегрессионные модели на основе декодера (например, семейство *GPT* [5]).

Такие модели, предобученные на корпусах текстов, научились генерировать связные, грамматически правильные и контекстуально уместные ответы. Категориальная модель системы с внешним контролем (см. рисунок 1) оперирует четырьмя ключевыми объектами:  $Q$  (исходный запрос),  $Q'$  (запрос, обогащенный синтаксическими аннотациями),  $A$  (черновой ответ генеративной модели) и  $A'$  (финальный верифицированный ответ).

Модуль постобработки представлен в модели морфизмом  $f_3: A \rightarrow A'$ . Такой подход позволяет сохранить наглядность категориального описания, фокусируясь только на входных и выходных состояниях системы.

*Funnel-Transformer* можно представить в виде двух частей: первая из них начинается с большого количества узлов, которое постепенно уменьшается, достигая самой узкой точки, после чего последующие слои вновь увеличивают количество узлов (см. рисунок 2): сначала происходит глобальное сжатие информации, затем добавляется локальная детализация. Эта архитектура обеспечивает баланс между скоростью обработки и точностью результата.

Для моделирования процесса предобработки данных применяется аппарат теории категорий [10, 11]. В данном подходе теория категорий служит для формального описания преобразований между различными представлениями данных, где основное внимание уде-

ляется морфизмам. В контексте обработки данных трансформером процесс перехода от исходных данных к подготовленным объектам формализуется через категорию  $C$ , объектами которой являются:

- $K_1$  – входные компоненты (слова, фразы, предложения);
- $K_2$  – правила формальных грамматик Хомского;
- $K_3$  – результаты предобработки.

Преобразования между этими объектами задаются морфизмами:

- $f_1: K_1 \rightarrow K_2$  – применение грамматических правил к входным данным;

- $f_2: K_2 \rightarrow K_3$  – преобразование синтаксических структур в векторные представления;
- $f_3 = f_2 \circ f_1: K_1 \rightarrow K_3$  – композиция преобразований морфизмов, удовлетворяющих условию коммутативности.

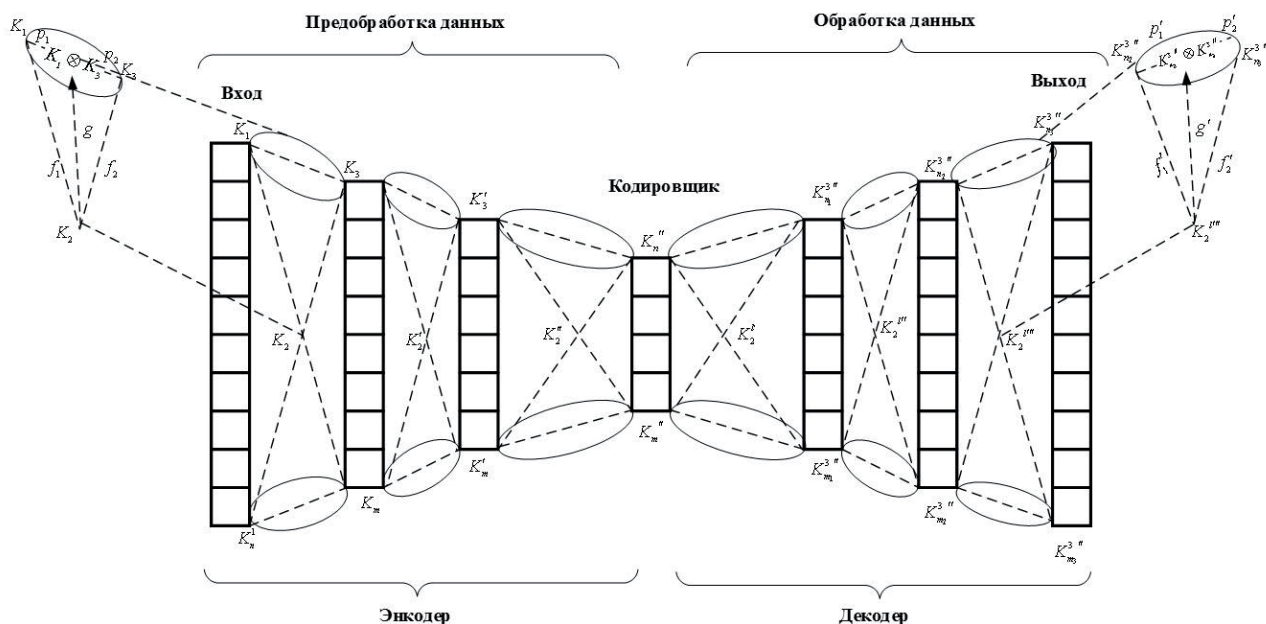


Рисунок 2 – Структура модели Funnel-Transformer

Данный процесс может быть представлен последовательностью коммутативных диаграмм, приведённых на рисунке 2. Процесс предобработки может быть описан через декартово произведение  $K_1 \times K_2$ , представляющее все возможные пары «входной компонент – грамматическое правило». Морфизм  $g: K_1 \times K_2 \rightarrow K_3$ , являющийся результатом процесса предобработки, обеспечивает переход к конечному представлению данных. Такой формализм позволяет:

- единообразно описывать этапы обработки данных;
- анализировать свойства композиции преобразований;
- выявлять инварианты архитектуры обработки.

В итоге переходы от одной диаграммы к следующей можно выражать через новую категорию  $C$ , включающую множество объектов  $O = \{K_1, K_2, K_3\}$ . Множество взаимодействий между этими категориями, внутри которых расположены информационные объекты, представимо коммутативным треугольником по категории и отдельным подкатегориям.

Категория  $K_2$  образует класс объектов с заданным отношением для любых пар  $(K_2, K_1)$  и  $(K_2, K_3)$ , отношения которых определены множеством морфизмов (связи между элементами объектов):  $\{f_1, f_2, p_1, p_2\}$ . Процессы предобработки данных по правилам  $K_2$  рассматриваются как взаимосвязанные процессы между входными компонентами  $K_1$  и результатом предобработки данных  $K_3$ . Объект, являющийся результатом такого взаимодействия, представляется декартовым произведением  $g: K_2 \rightarrow K_1 \otimes K_3$ , где  $\otimes$  – декартово произведение,  $g$  – единственный морфизм, связывающий результаты с процессом предобработки данных.

Процесс предобработки данных первого слоя нейронной сети соответствует отношениям, формализованным в виде морфизмов. Данные отношения обладают свойствами: результат взаимодействия информационных объектов может быть представлен декартовым произведением  $K_1 \otimes K_3$  с морфизмами  $p_1: K_1 \otimes K_3 \rightarrow K_1$  и  $p_2: K_3 \otimes K_1 \rightarrow K_3$  такими, что для любого объекта  $K_2$  с морфизмами  $f_1: K_2 \rightarrow K_1$  и  $f_2: K_2 \rightarrow K_3$  существует единственный морфизм  $g: K_2 \rightarrow K_1 \otimes K_3$ , при котором выделяется диаграмма в форме коммутативного треугольника.

Таким образом, обработка запроса формализована в терминах теории категорий, где объекты – ключевые представления данных в системе, морфизмы – преобразования между объектами, коммутативные диаграммы позволяют наглядно отследить соответствие между альтернативными путями обработки.

## 1.2 Физико-математический подход к обработке текстов

Физико-математический подход основывается на введении геометрических характеристик для анализа текстовых данных. Одним из центральных понятий здесь является про-

странство Минковского, где каждое слово или предложение представлено точкой в четырёхмерном пространстве.

Использование пространственно-временной размерности Минковского позволяет выстраивать пространственную конфигурацию, в которой каждая точка представляет собой событие, связанное с определённым входным компонентом (или результатом его предобработки). Такой подход обеспечивает построение набора данных на основе описания категориальных связей. Выстраивание пространственной конфигурации по Минковскому позволяет получить параллельное соотношение всех модулей и правил с использованием результатов предобработки в любой момент времени в виде набора данных, а выстраивание отношений с другими конусами позволяет определить новые правила, которые необходимы для предобработки следующих входных компонентов [12]. В таком пространстве каждому событию можно сопоставить точку, пространственные координаты которой описывают место, где событие произошло, а временная координата – временной компонент, когда оно случилось:

$K_1 = \{C_1, C_2, C_3, \dots, C_n\}$  – входные компоненты, представляющие множество вариантов запросов (текст, картинки, звуки, чертежи, алфавит и т.п.);

$K_2 = \{P_1, P_2, P_3, \dots, P_l\}$  – множество правил по формальным грамматикам Хомского;

$K_3 = \{R_1, R_2, R_3, \dots, R_m\}$  – множество вариантов ответов.

Здесь  $n, l, m$  – мощности соответствующих множеств.

При анализе моделей обработки ЕЯ целесообразно классифицировать их по классам задач (классификация текста, анализ семантического сходства, машинный перевод, обработка разговорного языка) для решения которых они предназначены [13, 14].

Обучение представлений подразумевает извлечение и преобразование дискретных или сложных объектов, например, слов, предложений, изображений, в числовые векторные представления фиксированной размерности, которые могут быть перенастроены под различные задачи. Это свидетельствует о том, что модель может адаптироваться к требованиям конкретной предметной области. В результате снижается потребность в обширных размеченных данных для каждой задачи, поскольку базовые репрезентации уже содержат ключевые аспекты языка и его использования. Модели становятся более универсальными, способными решать сложные задачи с высокой степенью точности, опираясь на предшествующий опыт.

### 1.3 Семиотические аспекты обработки текста

Задачи ЕЯ могут быть представлены семиотической системой связи [15, 16], которая представляет собой совокупность элементов, позволяющих передавать и интерпретировать информацию посредством символов, изображений, звуков, жестов и других знаковых форматов. В семиотических системах представлен обширный контекст, и результаты запросов могут быть разными, но сохранять основной смысл. При обратном запросе оригинальный текст не получится [17].

Семантическое пространство представляется формулой  $S=(W, V)$ , где:  $S$  – семантическое пространство,  $W$  – словарь,  $V$  – векторное пространство, в котором каждому слову сопоставляется вектор, отражающий его смысловые характеристики. Каждое слово  $w$  представляется точкой в пространстве Минковского с координатами  $(x, y, z, t)$ , где:  $x, y, z$  – пространственные координаты, зависящие от семантического смысла слова,  $t$  – временная координата, отражающая эпоху или контекст использования слова [12]. Пространство Минковского представляет собой четырёхмерное псевдоевклидово пространство-время, где метрика задаётся как:  $ds^2 = dt^2 - dx^2 - dy^2 - dz^2$ , где  $dx, dy, dz$  – разности пространственных координат (семантических измерений),  $dt$  – разность временной координаты (временного аспекта).

В контексте семантического пространства координаты  $x, y, z$  представляют семантическую близость слов (например, синонимия, родственные отношения, тематическая принадлежность) и зависят от семантических признаков. Чем ближе точки, тем более близкими по смыслу считаются слова, координата  $t$  отражает исторический контекст или временной период, в течение которого слово употреблялось или приобрело популярность. Большое значение  $t$  может означать, что слово старое или давно вышло из активного употребления.

Пусть имеется два слова  $w_1$  и  $w_2$  с координатами:  $w_1=(x_1,y_1,z_1,t_1)$  и  $w_2=(x_2,y_2,z_2,t_2)$ . Интервал между ними составит:  $ds^2(w_1,w_2)=(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2-(t_2-t_1)^2$ . Пространственные координаты кодируют инвариантные семантические признаки слова, временная координата является мерой его контекстуальной, диахронической или прагматической модификации. Особенностью метрики является её способность различать два типа семантических отношений:

- Смысловая близость (положительный  $ds^2$ ) достигается, когда совпадение по инвариантным семантическим признакам  $(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2$  перевешивает контекстуально-временной разрыв  $(t_2-t_1)^2$ . Это соответствует случаю, когда слова являются синонимами или принадлежат одному семантическому полю.
- Смысловой разрыв или сдвиг (отрицательный  $ds^2$ ) возникает, когда различия в контексте, эпохе или прагматике использования  $(t_2-t_1)^2$  становятся настолько значимыми, что преобладают над формальной семантической схожестью.

Метрика Минковского расширяет модель семантического пространства, вводя в него конфликтующее измерение. Это позволяет более адекватно отражать сложность и нелинейность семантических связей в ЕЯ.

Пусть дано семантическое пространство  $S$ , представленное как векторное пространство над полем вещественных чисел. Вводится метрика, аналогичная метрике Минковского, и применяется к векторам слов. Тогда расстояние между двумя словами  $w_1$  и  $w_2$  определяется следующим образом [9]:

$$d(w_1, w_2)^2 = t_{w_1} w_1^T A^{-1} t_{w_2} w_2 - s_{w_1}^T B s_{w_2} ,$$

где  $t_w$  – представляет временную компоненту слова  $w$ ,  $S_w$  – пространственный компонент (семантический вектор),  $A$  и  $B$  – положительно определённые матрицы, характеризующие влияние временной и пространственной компонент соответственно.

На рисунке 3 представлена структура объектов с правилами поиска и запроса для конкретного языка. Ответ осуществляется на языке запроса с возможностью перевода по лингвистическим правилам на другой язык [6, 18].

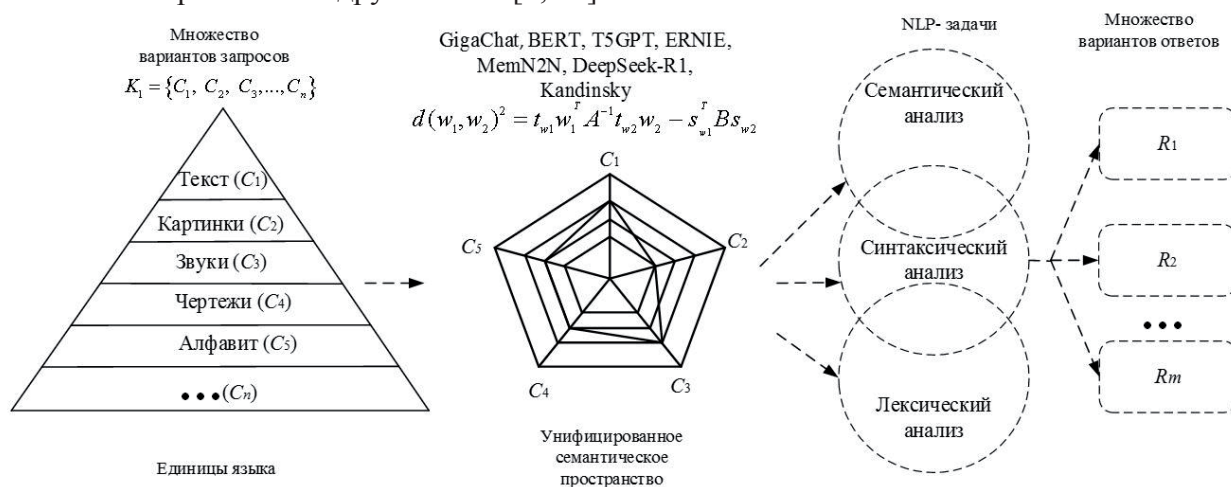


Рисунок 3 – Единое семантическое пространство для различных языковых структур

Исходный текст подаётся модели на начальном этапе обработки и преобразуется моделью в числовое представление в виде набора векторов, удобных для последующей обработки. Применяя ранее полученные знания с помощью механизма внимания, модель изменяет структуру полученного вектора таким образом, чтобы он соответствовал структуре текста на требуемом языке перевода. Итоговый вектор подвергается обратному преобразованию в читаемый текст на целевом языке. Модель выводит готовый перевод на целевом языке.

Формальные грамматики обеспечивают основу для анализа и генерации предложений, разбивая их на составляющие элементы и определяя правила их сочетания [19]. Грамматики Хомского могут применяться в контексте современных языковых моделей на основе трансформеров. Единицы языка состоят из компонентов:  $G = \{N, T, P, S\}$ , где:  $N$  – алфавит конкретных символов, формирующий конечный алфавит нетерминальных символов;  $T$  – конечный алфавит терминальных символов (совпадает с алфавитом языка, задаваемого грамматикой);  $P$  – правила вывода, морфология;  $S$  – начальный нетерминал грамматики  $G$ .

Граматики Хомского могут оказать существенную поддержку в снижении генерации недостоверных сведений.

#### 1.4 Модели трансформеров

Модели, такие как *BERT*, используют только компонент энкодера. Это позволяет им эффективно выполнять задачи, такие как текстовая классификация, распознавание именованных сущностей и ответы на вопросы [20].

Модели типа *GPT* функционируют на основе декодера, что позволяет им успешно справляться с задачами, связанными с продолжением текста, ведением диалогов и оценкой вероятности текстов.

Таким образом, *BERT* ориентирован на анализ и классификацию, *GPT* специализируется на генерации текста и взаимодействии [21, 22].

Недостатки известных моделей: неустойчивость (потеря качества при изменении данных); предвзятость (модель даёт ответы, не соответствующие этике); нестабильность вывода (получение разных результатов запроса).

## 2 Особенности и результаты тестирования диалоговой модели

Для создания диалоговой модели использована БЯМ ГигаЧат 2.0, предназначенная для работы с ЕЯ. Модель основана на архитектуре трансформеров, схожей с семейством *GPT*, и способна решать широкий спектр задач [20].

Для работы с различной длиной входных данных и сохранения контекста в продолжительных диалогах применяется механизм направленного внимания, реализованный через синтаксически-обогащённый запрос. Модель включает внешние модули проверки достоверности, взаимодействующие с дополнительными источниками данных для верификации фактов. Механизм внимания предлагается использовать на этапе предобработки данных [23]. Это упрощает и ускоряет обработку, но появляется риск потери деталей и возможного ухудшения качества на коротких текстах. Важно внимательно подбирать архитектуру и настраивать гиперпараметры для оптимального баланса между скоростью и точностью обработки.

Построение компактного представления текста осуществляется через каскад блоков компрессии, которые уменьшают длину последовательности, при сохранении существенных признаков. Это может быть достигнуто введением функции компрессии  $L(x)$  (см. рисунок 4). На каждом временном шаге  $t$  глобальная модель внимания вычисляет вектор веса выравнивания переменной длины  $\alpha_t$  на основе текущего состояния цели  $h_t$  и всех состояний источ-

ника  $h_s$  [24]. Глобальный вектор контекста  $c_t$  вычисляется как средневзвешенное значение, согласно  $\alpha_t$  по всем состояниям источника. Цель – получить вектор контекста на основе всех скрытых состояний кодирующей рекуррентной нейронной сети. Этот тип внимания охватывает всё пространство входных состояний [25, 26].

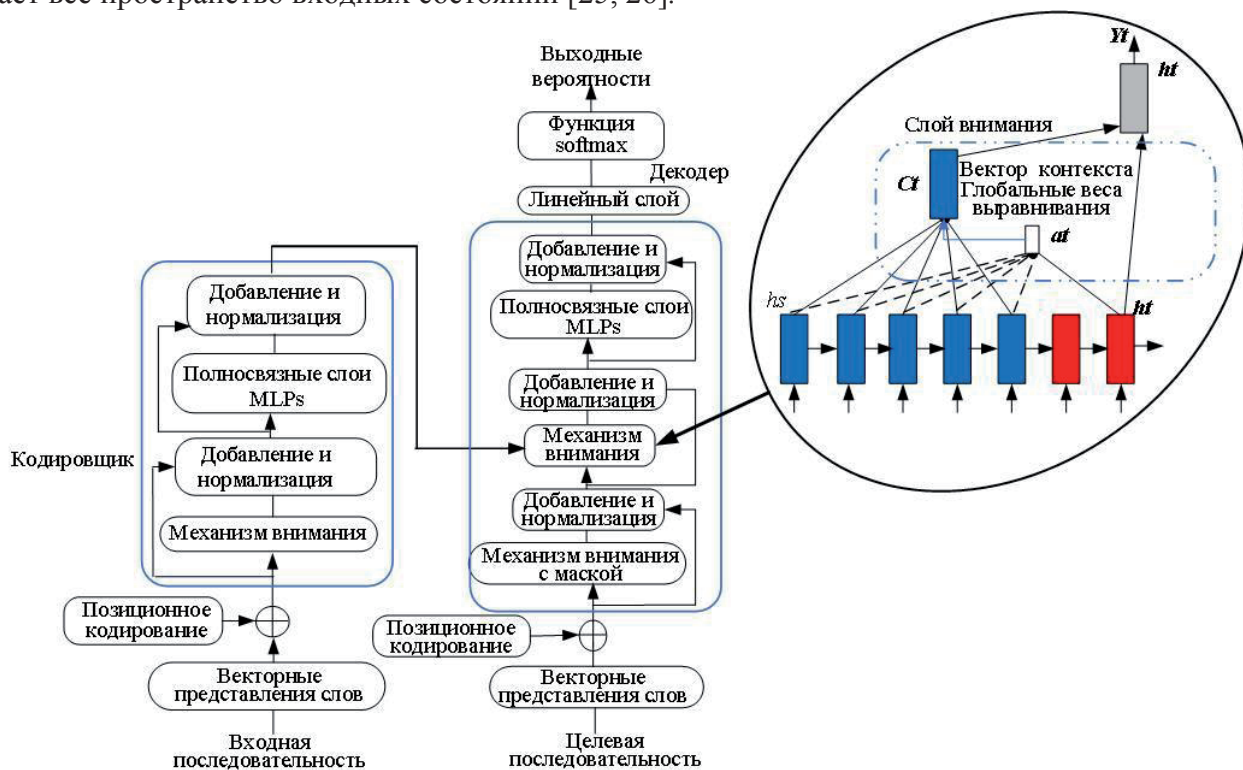


Рисунок 4 – Структура языковой модели с архитектурой трансформера

Алгоритм работы БЯМ включает: сбор данных; предобработку данных; работу диалоговой модели, имеющей архитектуру трансформера.

В ходе тестирования проверялись ключевые характеристики модели: качество генерации текста и правильность понимания вопросов; временные показатели откликов на запросы; надёжность предоставляемой информации.

Дообучение проведено на специализированном наборе данных объёмом около 15 000 пар «вопрос-ответ» из образовательной сферы (учебники по гуманитарным наукам, сборники задач). Для процесса дообучения модели использовался оптимизатор *AdamW* [27]. Для интеграции грамматических правил в процесс генерации использован механизм синтаксически-обусловленного внимания, который на этапе предобработки формирует синтаксические признаки размерностью 128 и использует их для направления фокуса модели на структурно корректные паттерны. Достоверность ответов определялась сочетанием трёх показателей: фактологическая точность – соответствие утверждений в ответе проверяемым фактам из авторитетных источников; логическая связность – отсутствие противоречий в ответе, соблюдение причинно-следственных связей и правил логики; синтаксическая и семантическая корректность – грамматическая правильность предложений и соответствие смысла запросу.

В качестве тестового использовался набор данных из 500 запросов трёх категорий: фактологические (200) – вопросы с однозначным проверяемым ответом; структурированные/логические (200) – математические задачи, задачи на логику, запросы на построение алгоритма; провоцирующие (100) – запросы, содержащие противоречия или отсылающие к несуществующим фактам (для оценки склонности к галлюцинациям).

Для оценки использовались следующие метрики: доля фактологически верных утверждений в ответе, доля вымышленных или неподтверждённых утверждений в ответе, оценка логической связности и последовательности ответа, среднее время генерации ответа. Результаты приведены в таблице 1.

Таблица 1 – Результаты экспериментального сравнения

Модель	Доля фактологически верных утверждений в ответе	Доля вымышленных или неподтверждённых утверждений в ответе	Оценка логической связности и последовательности ответа	Среднее время генерации ответа
Baseline (ГигаЧат 2.0)	78.5%	18.2%	3.8	0.45
Предложенная модель	91.3%	7.1%	4.4	0.68

Полученные результаты позволяют сопоставить предлагаемую модель с другими подходами (см. таблицу 2).

Таблица 2 – Качественное сопоставление подходов к обеспечению достоверности языковых моделей

Критерий / Метод	RLHF [4, 5]	RAG [5]	Нейро-символьные гибриды [6, 7]	Предлагаемый гибридный метод
Основной механизм	Смещение выхода через предпочтения	Поиск по внешним БЗ	Интеграция нейросетей с логич. правилами	Тройная интеграция: грамматики + категории + геом. модель
Фактологическая точность	Косвенно, через обратную связь	Высокая (за счёт БЗ)	Зависит от системы правил	Высокая (контроль на входе/выходе + RAG-верификация)
Синтаксический/ структурный контроль	Отсутствует	Отсутствует	Частичный (зависит от грамматик)	Высокий (формальные грамматики Хомского)
Объяснимость	Низкая («чёрный ящик»)	Средняя (можно указать источник)	Высокая (логический вывод)	Высокая (категориальная модель потоков)
Учёт контекста/связность	Высокий (за счёт обучения)	Ограничен контекстом запроса	Жёсткий (правила)	Высокий (геом. модель Минковского + генеративная БЯМ)
Вычислительная сложность	Очень высокая (обучение с подкр.)	Высокая (поиск + генерация)	Средняя-высокая	Средняя (увеличение ~50% к baseline, см. таблица 1)
Ключевое ограничение	Затраты на разметку, галлюцинации	Качество БЗ, связность текста	Жёсткость, сложность интеграции	Время отклика, творческие задачи

В отличие от *RLHF* и *RAG* предлагаемый метод обеспечивает системное преимущество в задачах, требующих одновременно высокой структурной корректности, фактологической точности и объяснимости. По вычислительной сложности метод занимает промежуточное положение: он легче масштабируемого сбора человеческих предпочтений (*RLHF*), но требует больше ресурсов, чем *RAG*. Таким образом, метод заполняет нишу проектирования контролируемых и верифицируемых диалоговых систем для предметно-ориентированных областей, где цена ошибки высока, а требования к корректности преобладают над требованиями к минимальной задержке.

Добавление модуля на основе грамматик Хомского даёт значительный прирост точности и снижает уровень галлюцинаций. Это подтверждает гипотезу о важности синтаксической

фильтрации. Улучшение качества сопровождается увеличением времени ответа (на ~50% с 0.45 до 0.68 с).

Эффективность предложенного подхода снижается в следующих условиях.

- При работе с творческими задачами (написание эссе, генерация художественных текстов), где строгий синтаксический контроль может ограничивать творчество.
- В условиях неопределённых или противоречивых запросов, когда грамматический анализ не позволяет однозначно классифицировать намерение пользователя.
- При обработке специализированной терминологии, не охваченной правилами грамматик Хомского.
- В сценариях реального времени с жёсткими требованиями к задержке, поскольку дополнительный модуль предобработки увеличивает время отклика на 15-20%.

Наибольшая эффективность достигается при работе со структурированными запросами: решение математических задач, проверка грамматики, фактологический поиск, технические консультации.

## Заключение

Основными результатами работы являются следующие:

- разработана гибридная архитектура диалоговой модели на основе ГигаЧат 2.0 и интегрирующая формальные грамматики Хомского для синтаксического контроля на входе и выходе;
- предложено формальное описание работы диалоговой модели на языке теории категорий, что упрощает анализ потоков данных;
- осуществлено моделирование семантики в пространстве Минковского для отделения инвариантного значения языковых единиц от их контекстуально-временной динамики, что улучшает логическую связность ответов;
- предложенный гибридный метод может снижать уровень галлюцинаций и повышать фактологическую точность и логическую связность ответов.

## Список источников

- [1] **Soydaner D.** Attention mechanism in neural networks: where it comes and where it goes. *Neural Computing and Applications*. 2022. Vol. 34, Issue 16. P.13371–13385. DOI: 10.1007/s00521-022-07366-3.
- [2] **Dai Z., Lai G., Yang Y., Le Q.V.** Funnel-Transformer: Filtering out Sequential Redundancy for Efficient Language Processing // arXiv:2006.03236. 2020. DOI:10.48550/arXiv.2006.03236.
- [3] **Elkhatat A.M.** Evaluating the authenticity of ГигаЧат responses: a study on text-matching capabilities. *Int J Educ Integr*. 2023. Vol.15. P.1-23. DOI:10.1007/s40979-023-00137-0.
- [4] **Bai Y. et al.** Constitutional AI: Harmlessness from AI Feedback // arXiv:2212.08073. 2022. DOI: 10.48550/arXiv.2212.08073.
- [5] **Brown T.B. et al.** Language Models are Few-Shot Learners // arXiv:2005.14165. 2020. DOI: 10.48550/arXiv.2005.14165.
- [6] **Ахо А., Ульман Дж.** Теория синтаксического анализа, перевода и компиляции. Т. 2: Синтаксический анализ. М.: Книга по требованию, 2012. 487 с.
- [7] **Clark P., Tafford O., Richardson K.** Transformers as Soft Reasoners over Language. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)*. 2020. P.3882–3890. DOI: 10.24963/ijcai.2020/53
- [8] **Шавгулидзе Н.Е.** Размерность Минковского и фракталы в р-адических пространствах // Обучение фрактальной геометрии и информатике в вузе и школе в свете идей академика А.Н. Колмогорова: материалы XVI Колмогоровских чтений: 3-й Международной научно-методической конференции, Кострома, 07–09 декабря 2021 года. Кострома: Костромской государственный университет, 2021. С.82-87.
- [9] **Vaswani A., Shazeer N., Parmar N. et al.** Attention Is All You Need. *Advances in Neural Information Processing Systems* 30 (NIPS 2017). 2017. P.5998–6008. <https://arxiv.org/abs/1706.03762>. DOI: 10.48550/arXiv.1706.03762.
- [10] **Маклейн С.** Категории для работающего математика / Пер. с англ. В.А. Артамонова, под ред. В.А. Артамонова. 2-е изд. М.: Физматлит, 2004. 352 с.

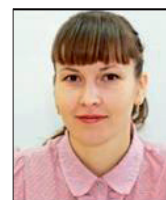
- [11] **Антонов В.В., Куликов Г.Г., Кромина Л.А., Родионова Л.Е.** Концепция программно-аналитического комплекса образовательного процесса на основе онтологии и искусственных нейронных сетей. *Онтология проектирования*. 2021. Т.11, №3(41). С.339-350. DOI:10.18287/2223-9537-2021-11-3-339-350.
- [12] **Vlontzos A., Rocha H.B., Rueckert D., Kainz B.** Causal Future Prediction in a Minkowski Space-Time // arXiv.2008.09154. 2020. DOI: 10.48550/arXiv.2008.09154.
- [13] **Arsalan M., Raza S.A., Singh A.P.** Transformers in Natural Language Processing: A Comprehensive Review. *International Journal for Research in Applied Science and Engineering Technology*. 2024. Vol.12, Issue 5. P.5591-5597. DOI :10.22214/ijraset.2024.62863.
- [14] **Сологуб Г.Б., Пухов В.А.** Проблемы классификации текстов естественного языка методами классического машинного обучения. *Моделирование и анализ данных*. 2023. Т.13 №2. С.64–76. DOI: 10.17759/mda.2023130203.
- [15] **Choi D., Spangner L., Hidey C., Grabowski P., Eskander R.** Revisiting Funnel Transformers for Modern LLM Architectures with Comprehensive Ablations in Training and Inference Configurations // arXiv.2504.02877. 2025. DOI: 10.48550/arXiv.2504.02877.
- [16] **Санников С.В.** Семиотические аспекты лингвокультурного трансфера: к проблеме типологии формул передачи информации. *Universum: Филология и искусствознание. Электронный научный журнал*. 2018. №9 (55). <https://7universum.com/ru/philology/archive/item/6341>.
- [17] **Валькман Ю.Р., Тарасов В.Б.** От онтологий проектирования к когнитивной семиотике // *Онтология проектирования*. 2018. Т.8, №1(27). С.8–34. DOI: 10.18287/2223-9537-2018-8-1-8-34.
- [18] **Morris C.W.** Foundations of the Theory of Signs. The Hague: Mouton, 1971. (Original 1938). DOI: 10.1515/css-2019-0001
- [19] **Хомский Н.** Язык и проблема знания. *Вестник Московского университета*. Сборник научных трудов. 1996. Выпуск 6. С.157–185.
- [20] **Аксьёнов К.А., Сунь Л.** Эволюция и современное состояние систем ответов на вопросы: технологии распознавания намерений и именованных сущностей с использованием модели BERT. *Инженерный вестник Дона*. 2024. № 7(115). С.761-773.
- [21] **Евстигнеев М.Н.** Планирование учебного занятия по иностранному языку с помощью технологий генеративного искусственного интеллекта. *Вестник Тамбовского университета. Серия: Гуманитарные науки*. 2024. Т.29, №3. С.617-634. DOI: 10.20310/1810-0201-2024-29-3-617-634.
- [22] **Hochreiter S.** The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 1998. Vol.6(2). P.107–116. DOI: 10.1142/S0218488598000094.
- [23] **Bahdanau D., Cho K., Bengio Y.** Neural Machine Translation by Jointly Learning to Align and Translate // arXiv.1409.0473. 2015. P.1-15. DOI: 10.48550/arXiv.1409.0473.
- [24] **Luong M.T., Pham H., Manning C.D.** Effective Approaches to Attention-based Neural Machine Translation // arXiv:1508.04025. 2015. DOI: 10.48550/arXiv.1508.04025.
- [25] **Kardakis S., Perikos I., Grivokostopoulou F., Hatzilygeroudis I.** Examining Attention Mechanisms in Deep Learning Models for Sentiment Analysis. *Applied Sciences*. 2021. Vol.11, Issue 9(3883). DOI: 10.3390/app11093883.
- [26] **Топольян С.Г.** Развитие генеративных моделей в логических системах принятия решений. *Вестник науки*. 2025. №5(86). С.671-678.
- [27] **Loschilov I., Hutter F.** Decoupled Weight Decay Regularization // arXiv:1711.05101. 2017. DOI: 10.48550/arXiv.1711.05101.

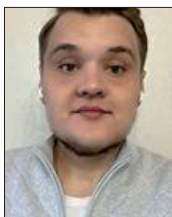
## Сведения об авторах



**Антонов Вячеслав Викторович**, 1956 г. рождения. Окончил Башкирский государственный университет (1979), к.т.н. (2007), д.т.н. (2015). Заведующий кафедрой автоматизированных систем управления УУНИТ. В списке научных трудов более 200 работ в области построения интеллектуальных систем. AuthorID (РИНЦ): 530537. AuthorID (Scopus): 57200254522; ResearcherID (WoS): AАН-5121-2019. [antonov.v@bashkortostan.ru](mailto:antonov.v@bashkortostan.ru).

**Родионова Людмила Евгеньевна**, 1984 г. рождения. Окончила УГАТУ (2007), к.т.н. (2019). Доцент кафедры автоматизированных систем управления УУНИТ. В списке научных трудов около 30 работ в области проектирования программных аналитических комплексов. AuthorID (РИНЦ): 852968. Author ID (Scopus): 57221335881 ORCID: 0000-0003-4041-0365; Researcher ID (WoS): AAU-3498-2020. [lurik@mail.ru](mailto:lurik@mail.ru). ✉





**Пальчевский Евгений Владимирович**, 1994 г. рождения. Окончил УГАТУ (2019). Кандидат технических наук (2024), доцент кафедры индустриального программирования РТУ МИРЭА. В списке научных трудов около 200 работ в области разработки программного обеспечения и интеллектуальных вычислений. AuthorID (РИНЦ): 837544. Author ID (Scopus): 57220744490; Researcher ID (WoS): ABB-2403-2021. *teelp@inbox.ru*.



**Суворова Вероника Александровна**, 1982 г. рождения. Окончила УГАТУ (2004), к.т.н. (2012). Доцент кафедры автоматизированных систем управления УУНиТ. В списке научных трудов около 30 работ в области управления организационными системами. AuthorID (РИНЦ): 541074. Author ID (Scopus): 57263144400 ORCID 0009-0009-6005-0695. *milana\_da@mail.ru*.

Поступила в редакцию 07.07.2025, после рецензирования 25.12.2025. Принята к публикации 15.01.2026.



Scientific article

DOI: DOI:10.18287/2223-9537-2026-16-1-125-138

## Enhancing the reliability of dialogue model responses using formal grammars and category theory

© 2026, V.V. Antonov<sup>1</sup>, L.E. Rodionova<sup>1</sup>✉, E.V. Palchevsky<sup>2</sup>, V.A. Suvorova<sup>1</sup>

<sup>1</sup>Ufa University of Science and Technology, Ufa, Russia

<sup>2</sup>MIREA – Russian Technological University (RTU MIREA), Moscow, Russia

### Abstract

This study is devoted to enhancing the reliability of dialog model responses in tasks requiring high accuracy. A hybrid approach is proposed that combines fine-tuning the model on domain-specific data with the integration of Chomsky formal grammars into the query processing pipeline. An architecture is introduced that unifies formal syntactic control with the generative capabilities of the fine-tuned GigaChat 2.0 model. The query processing workflow incorporates a module for preliminary analysis and filtering of user queries based on Chomsky context-free grammars. To align the operation of the generative model with the identified syntactic structure, a two-level control mechanism is proposed. At the input stage, a syntactically enriched query guides the generation process toward appropriate structural query patterns. At the output stage, the verification module performs selection and correction of the generated response based on formal criteria. To enhance explainability and enable formal analysis of data flows, category theory is employed, in which queries, intermediate representations, and responses are modeled as objects and morphisms of a category. To improve logical coherence and contextual awareness, a geometric representation of semantics in Minkowski space is proposed, whose metric makes it possible to separate the invariant meaning of a word from its contextual and temporal dynamics. Experimental evaluation on a specialized set of 500 factual and structured queries demonstrates that the proposed method reduces the generation of fabricated statements by 22% compared to the baseline GigaChat 2.0 model. The results show that the combined use of formal grammars, categorical formalism, and a geometric model of semantics enhances the reliability of responses in domain-specific dialogues.

**Keywords:** *response reliability, dialogue model, transformer, Chomsky formal grammars, category theory, Minkowski space, semantic space.*

**For citation:** Antonov VV, Rodionova LE, Palchevsky EV, Suvorova VA. Enhancing the reliability of dialogue model responses using formal grammars and category theory [In Russian]. *Ontology of Designing*. 2026; 16(1): 125-138. DOI:10.18287/2223-9537-2026-16-1-125-138.

**Financial Support:** The research was financially supported by the Ministry of Science and Higher Education of the Russian Federation of the state assignment to higher education institutions No. FRRR-2026-0006.

**Authors' contributions:** Antonov V.V. – development of the article structure and selection of sources; Rodionova L.E. – preparation of examples; Palchevsky E.V. – development of the dialogue model; V.A. Suvorova – verification of results.

**Conflict of interest:** The authors declare no conflict of interest.

## List of figures and tables

Figure 1 – Architecture of the hybrid system with external control

Figure 2 – Structure of the Funnel-Transformer model

Figure 3 – Unified semantic space for various linguistic structures

Figure 4 – Structure of a transformer-based language model

Table 1 – Results of the experimental comparison

Table 2 – Qualitative comparison of approaches to ensuring the reliability of language models

## References

- [1] **Soydaner D.** Attention Mechanism in Neural Networks: Where It Comes from and Where It Goes. *Neural Computing and Applications*. 2022; 34(16): 13371–13385. DOI: 10.1007/s00521-022-07366-3.
- [2] **Dai Z, Lai G, Yang Y, Le QV.** Funnel-Transformer: Filtering out Sequential Redundancy for Efficient Language Processing // arXiv: 2006.03236.[cs.CL]. 2020. DOI: 10.48550/arXiv.2006.03236.
- [3] **Elkhatat AM.** Evaluating the authenticity of ГигаЧат responses: a study on text-matching capabilities. *International Journal for Educational Integrity*. 2023; 15: 1-23. DOI: 10.1007/s40979-023-00137-0.
- [4] **Bai Y. et al.** Constitutional AI: Harmlessness from AI Feedback // arXiv:2212.08073 [cs.AI]. 2022. DOI: 10.48550/arXiv.2212.08073.
- [5] **Brown TB. et al.** Language Models are Few-Shot Learners // arXiv:2005.14165 [cs.CL]. 2020. DOI: 10.48550/arXiv.2005.14165.
- [6] **Aho A, Ullman J.** Syntax Analysis (Principles of Compiler Design, Vol. 2). Moscow: Kniga po Trebovaniyu, 2012. 487 p. (Original work published 1978). ISBN: 978-5-458-27407-4.
- [7] **Clark P, Tafford O, Richardson K.** Transformers as Soft Reasoners over Language. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)*. 2020. P.3882–3890. DOI: 10.24963/ijcai.2020/53.
- [8] **Shavgulidze NE.** Minkowski Dimension and Fractals in p-adic Spaces [In Russian]. *Materials of the XVIth Kolmogorov Readings: International Scientific-Methodical Conference*, Kostroma, December 7–9, 2021. Kostroma: Kostroma State University Press, 2021, P.82–87.
- [9] **Vaswani A, Shazeer N, Parmar N. et al.** Attention Is All You Need. *Advances in Neural Information Processing Systems* 30 (NIPS 2017). 2017. P.5998–6008. <https://arxiv.org/abs/1706.03762>. DOI: 10.48550/arXiv.1706.03762.
- [10] **Mac Lane S.** Categories for the Working Mathematician [In Russian]. (2nd ed., translated by V.A. Artamonov, edited by V.A. Artamonov). Moscow: Fizmatlit. 2004. 352 p.
- [11] **Antonov VV, Kulikov GG, Kromina LA, Rodionova LE.** The concept of a software and analytical complex of the educational process based on ontology and artificial neural networks [In Russian]. *Ontology of Designing*. 2021; 11(3): 339–350. DOI: 10.18287/2223-9537-2021-11-3-339-350.
- [12] **Vlontzos A, Rocha HB, Rueckert D, Kainz B.** Causal Future Prediction in a Minkowski Space-Time // arXiv.2008.09154 [cs.CV]. 2020. DOI: 10.48550/arXiv.2008.09154.
- [13] **Arsalan M, Syed Azeem Raza, Er. Aayush Pratap Singh.** Transformers in Natural Language Processing: A Comprehensive Review. *International Journal for Research in Applied Science and Engineering Technology*. – 2024; 12(5): 5591-5597. DOI: 10.22214/ijraset.2024.62863.
- [14] **Sologub GB, Pukhov VA.** Problems of classification of natural language texts by classical machine learning methods. *Data Modeling and Analysis*. 2023; 13(2): 64–76. DOI: 10.17759/mda.2023130203.
- [15] **Choi D, Spangner L, Hidey C, Grabowski P, Eskander R.** Revisiting Funnel Transformers for Modern LLM Architectures with Comprehensive Ablations in Training and Inference Configurations // arXiv.2504.02877 [cs.LG]. 2025. DOI: 10.48550/arXiv.2504.02877.
- [16] **Sannikov SV.** Semiotic Aspects of Linguocultural Transfer: Towards the Typology of Information Transfer Formulas [In Russian]. *Universum: Philology and Arts Studies. Online scientific journal*. 2018; 9(55). <https://7universum.com/ru/philology/archive/item/6341>.
- [17] **Valkman Yu R, Tarasov VB.** From ontologies of designing to cognitive semiotics [In Russian]. *Ontology of Designing*. 2018; 8(1): 8–34. DOI: 10.18287/2223-9537-2018-8-1-8-34.
- [18] **Morris CW.** Foundations of the Theory of Signs. The Hague: Mouton, 1971. (Original 1938). DOI: 10.1515/css-2019-0001.

- [19] **Chomsky N.** Language and the Problem of Knowledge [In Russian]. *Bulletin of Moscow State University. Collected Scientific Papers.* 1996; 6: 157–185.
  - [20] **Aksyonov KA, Sun L.** Evolution and current status of question answering systems: Technologies for Recognizing Intentions and Named Entities Using BERT Model [In Russian]. *Engineering Bulletin of the Don.* 2024; 7(115): 761–773.
  - [21] **Yevstigneev MN.** Planning a Lesson in a Foreign Language Using Generative AI Technologies [In Russian]. *Tambov State University Bulletin.* 2024; 29(3): 617–634. DOI: 10.20310/1810-0201-2024-29-3-617-634.
  - [22] **Hochreiter S.** Disappearing Gradient Problems in Recurrent Neural Network Training and Their Solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems.* 1998; 6: 107–116. DOI: 10.1142/S0218488598000094.
  - [23] **Bahdanau D, Cho K, Bengio Y.** Neural Machine Translation by Jointly Learning to Align and Translate // arXiv:1409.0473 [cs.CL].2015. DOI: 10.48550/arXiv.1409.0473.
  - [24] **Luong MT, Pham H, Manning CD.** Effective Approaches to Attention-based Neural Machine Translation // arXiv:1508.04025 [cs.CL].2015. DOI: 10.48550/arXiv.1508.04025.
  - [25] **Kardakis S, Perikos I, Grivokostopoulou F, Hatzilygeroudis I.** Examination of Attention Mechanisms in Deep Learning Models for Sentiment Analysis. *Applied Sciences.* 2021; 11(9): 3883. DOI: 10.3390/app11093883.
  - [26] **Topolyan SG.** Development of Generative Models in Logical Decision-Making Systems [In Russian]. *Bulletin of Science.* 2025; 5(86): 671–678.
  - [27] **Loschilov I, Hutter F.** Decoupled Weight Decay Regularization // arXiv:1711.05101 [cs.LG].2017. DOI: 10.48550/arXiv.1711.05101.
- 

## About the authors

**Vyacheslav Viktorovich Antonov** (b. 1956) graduated from Bashkir State University (1979), Candidate of Technical Sciences (2007), Doctor of Technical Sciences (2015). Head of the Department of Automated Control System, Ufa University of Science and Technology. The list of scientific papers contains more than 200 works in the field of intelligent systems development. AuthorID (RCI): 530537. Author ID (Scopus): 57200254522; Researcher ID (WoS): AAH-5121-2019. [antonov.v@bashkortostan.ru](mailto:antonov.v@bashkortostan.ru).

**Lyudmila Evgenievna Rodionova** (b. 1984) graduated from Ufa State Aviation Technical University (2007), Candidate of Technical Sciences (2019). Associate Professor, Department of Automated Control Systems, Ufa University of Science and Technology. The list of scientific papers includes about 10 works in the field of designing software analytical systems. AuthorID (RCI): 852968. Author ID (Scopus) 57221335881; Researcher ID (WoS): AAU-3498-2020. [lurik@mail.ru](mailto:lurik@mail.ru) ✉

**Evgeny Vladimirovich Palchevsky** (b. 1994) graduated from Ufa State Aviation Technical University (2019). Candidate of Technical Sciences (2024), Associate Professor, Department of Industrial Programming, RTU MIREA. The list of scientific publications includes about 200 works in the field of software development and intelligent computing. AuthorID (RSCI): 837544. Author ID (Scopus): 57220744490; ORCID 0000-0001-9033-5741; Researcher ID (WoS): ABB-2403-2021. [teelxp@inbox.ru](mailto:teelxp@inbox.ru).

**Veronika Aleksandrovna Suvorova** (b. 1982) graduated from Ufa State Aviation Technical University (2004), Candidate of Technical Sciences (2012). Associate Professor, Department of Automated Control Systems, Ufa University of Science and Technology. The list of scientific publications include approximately 30 works in the field of organizational systems management. AuthorID (RSCI): 541074. Author ID (Scopus): 57263144400; ORCID 0009-0009-6005-0695. [milana\\_da@mail.ru](mailto:milana_da@mail.ru).

---

Received July 7, 2025. Revised December 25, 2025. Accepted January 15, 2026.

---