УДК 004.891.2

ПРИМЕНЕНИЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ В СИСТЕМАХ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ. Ч. 1. Модели объяснения и большие языковые модели

А. А. Кулинич

Институт проблем управления им. В. А. Трапезникова РАН, г. Москва

✉ alexkul@rambler.ru

Аннотация. Большие языковые модели оказывают существенное влияние на многие сферы человеческой жизни: на образование, творчество, науку, бизнес. Исследуется применение больших языковых моделей для объяснения альтернатив решений, полученных системой поддержки принятия решений в условиях неопределенности. Рассмотрены классические и прагматические модели объяснения, предложенные философами. Сформулированы цели и задачи объяснения в процессах поддержки принятия решений в условиях неопределенности. Приведены концептуальный анализ функционирования больших языковых моделей и оценка их текущих способностей при решении типовых тестовых задач. Рассмотрены основные техники промптинга (системы запросов к языковой модели), позволяющие настроить языковую модель на генерацию объяснений для альтернатив решений конкретных проблем в предметной области. Рассмотрены техники промптов для поддержки прагматических и классических теорий объяснения альтернатив решений.

Ключевые слова: поддержка принятия решений, модели объяснения, цель объяснения, задачи объяснения, большая языковая модель, техники промптов.

ВВЕДЕНИЕ

Для поддержки принятия решений в социальных, политических, экономических и организационных системах в условиях неопределенности применяют «мягкий» системный анализ. Он основывается на принципе ограниченной рациональности [1], который утверждает, что в силу ограниченных когнитивных ресурсов лиц, создающих модель объекта, можно сформировать упрощенную системную модель с гипотетической структурой, параметры которой могут иметь лингвистические значения [2]. Построение модели в рамках «мягкого» системного анализа основано на изучении литературы, экспертных оценках, т. е. исходными данными являются неструктурированные данные – свободный текст. Модель описывает объект на ограниченном естественном языке, который предлагает эксперт.

Известным математическим аппаратом «мягкого» системного анализа являются когнитивные карты [3, 4]. Результаты моделирования в такой модели представлены на экспертном ограниченном естественном языке и интерпретируются самим экспертом в терминах его знаний, которых может быть недостаточно для получения нового решения. Интерпретация должна включать знания о предметной области, релевантные решению, но находящиеся за пределами ограниченного экспертного языка и упрощенной модели ситуации. Интерпретируя результаты моделирования, эксперт пытается объяснить и связать модельные процессы с процессами реального мира. В модели ситуации для объяснения результата формально могут быть построены причинно-следственные цепочки вывода результата [5]. Однако эти цепочки рассуждений выражены на ограниченном естественном языке и могут гарантировать строгость логического вывода в рамках принятых допущений и ограничений, но не возможность применения решения в реальном мире. Реальный мир намного разнообразнее и богаче.

В психологии считается, что принятие решений осуществляется интеллектом человека в психической среде, которая называется ментальным пространством. В ментальном пространстве отражены знания, элементы жизненного опыта эксперта, полученные на протяжении всей его жизни. В ментальном пространстве осуществляются мысленные операции – рассуждения, обобщения, интерпретация, объяснение и оценка возможных вариантов решений. Поддержать эти ментальные операции процессов принятия решений в условиях неопределенности можно с помощью систем искусственного интеллекта, в частности больших языковых моделей (англ. *Large Language Model,* LLM).

В настоящее время количество и разнообразие больших языковых моделей стремительно растет. Широкую известность и популярность в практическом применении получили такие большие языковые модели, как GPT-2 [6], GPT-3[7], InstructGPT [8] и GPT-4 [9] от компании OpenAI, BERT [10], RoBERTa, ALBERT [11] (разработчик – Google DeepMind). Активно используются разработки компании Meta AI*, такие как LLaMA (Large Language Model Meta AI) [12]. Очень популярна китайская языковая модель DeepSeek [13]. В России широко используют отечественные большие языковые модели GigaChat-2 от Сбера [14] и Yandex GPT 5.1 от компании Яндекс [15]. Подробный обзор текущего состояния разработок в области больших языковых моделей можно найти в работе [16].

Большие языковые модели применяются для решения самых разных задач в экономике, социальной сфере, медицине, образовании, государственном и муниципальном управлении и др. Приведем некоторые обзоры применения таких моделей: в научных исследованиях [17], для прогнозирования сложных экономических систем в современных условиях [18], в сфере государственного и муниципального управления [19], в области здравоохранения [20], в организациях и банковской сфере [21, 22], в качестве цифровых помощников [23], в бизнес-аналитике и принятии решений [24].

В работе [25] предложена методология использования больших языковых моделей для решения задач государственного и муниципального управления для реферирования и генерации текстового контента. Стратегическое направление в области цифровой трансформации государственного управления в Российской Федерации [26] определяет необходимость внедрения в сферу государственного управления систем искусственного интеллекта (ИИ), больших данных, Интернета вещей.

В рамках Форсайта по приоритетным направлениям фундаментальных и поисковых исследований в сфере ИИ [27] были выделены следующие приоритетные направления поисковых исследований в этой области: архитектуры и алгоритмы машинного обучения, вычисления и данные для ИИ, фундаментальные и генеративные модели, взаимодействие человека и ИИ, прикладные исследования для науки, образования и социальной сферы.

В центре внимания настоящей статьи находятся актуальные вопросы взаимодействия человека и систем искусственного интеллекта в процессах принятия решений в условиях неопределенности.

Цель статьи заключается в разработке методов и подходов (моделей и алгоритмов) поддержки принятия решений в условиях неопределенности с использованием больших языковых моделей.

Основные задачи, которые необходимо решить для достижения цели, видятся следующим образом.

Вначале необходимо понять, являются ли тексты объяснений, сгенерированные языковой моделью, объяснениями в смысле философских теорий объяснения, которые представляют собой фундаментальное основание методологии науки и научных исследований. Для этого проведем анализ существующих моделей объяснения и выясним их суть.

Вторая задача – это задача настройки большой языковой модели на генерацию корректных объяснений ситуаций для решения задач поддержки принятия решений в условиях неопределенности в тандеме с лицом, принимающим решение (ЛПР).

Третья задача – это оценка качества системы поддержки принятия решений, включающей ЛПР и большую языковую модель в качестве ассистента.

Статья состоит из двух частей. В первой части описываются первые две задачи. В частности, выделены три класса философских моделей объяснения, принятых в методологии науки. Выделенные классы моделей объяснения позволяют понять, является ли текст, сгенерированный большой языковой моделью, объяснением с научной точки зрения, и классифицировать эти тексты в выделенные классы моделей объяснения. Здесь также определены роль, цель и задачи объяснения в процессах принятия решений в условиях неопределенности. Приводятся сведения о текущей производительности больших языковых моделей ведущих иностранных и российских производителей. Это позволяет судить о возможности применения языковых моделей в качестве ассистента ЛПР в процес-

* Компания Meta Platforms, Inc. признана в России экстремистской, ее деятельность запрещена на территории Российской Федерации.

сах принятия решений. Определены возможности настройки большой языковой модели с помощью техник промптов (подсказок языковой модели) на генерацию объяснений в терминах выделенных классов моделей объяснения и задач объяснения, решаемых в условиях неопределенности.

Цель первой части статьи заключается в том, чтобы дать основные понятия и определения, которые будут использованы во второй части. Эта часть будет также полезна тем, кто впервые попытается применить большую языковую модель в своей системе поддержки принятия решений.

Во второй части статьи рассматривается вопрос измерения качества когнитивной системы поддержки принятия решений, включающей человека (ЛПР) и языковую модель. Для определения качества такой системы нужно измерить латентную переменную – удовлетворенность ЛПР объяснением языковой модели. Сформулированы критерии оценки степени удовлетворенности ЛПР при решении задач объяснения, обозначенных в первой части статьи. Предложена модель разумного респондента. Приводится пример оценки объяснений с использованием предложенной модели разумного респондента для двух российских больших языковых моделей (модель Сбера GigaChat 2.0 и модель Яндекса Yandex GPT 5 Pro).

1. МОДЕЛИ ОБЪЯСНЕНИЯ В МЕТОДОЛОГИИ НАУКИ

В философских, социологических, математических словарях и энциклопедиях можно найти различные определения объяснения, начиная от его свойств, структуры и заканчивая его ролью в научном познании. Приведем некоторые обобщенные определения.

В философской энциклопедии [28] объяснение – это рассуждение, посылки которого содержат информацию, достаточную для выведения из нее описания объясняемого явления. Объяснение представляет собой ответ на вопрос: почему данное явление происходит?

Функция объяснения определяется как функция научного познания, раскрытие сущности изучаемого объекта; она осуществляется посредством постижения закона, которому подчиняется данный объект, либо путем установления тех связей и отношений, которые определяют его существенные черты [29].

В методологии науки объяснение – это познавательная процедура, направленная на обогащение и углубление знаний о явлениях реального мира посредством включения этих явлений в структуру определенных связей, отношений и зависимостей, дающих возможность раскрыть существенные черты данного явления [30].

В философии и методологии науки ставится задача создания единой универсальной теории объяснения, применимой в различных областях человеческой деятельности, например, в ходе получения знаний в рамках таких научных дисциплин, как физика, химия, биология, социология и др. Вариантами создаваемой общей теории объяснения являются различные модели объяснения, предложенные разными исследователями или коллективами исследователей. Разрабатываемые модели объяснения основываются на многочисленных примерах созданных теорий реального мира и в целом эти модели отражают механизмы работы человеческого интеллекта в процессах исследования, а также понимания и обоснования эмпирических наблюдений и фактов [31].

Важно понять, насколько искусственные большие языковые модели способны объяснять реальные объекты или события в процессах принятия решений в условиях неопределенности.

Необходимо отметить, что в настоящее время есть множество тестовых программ (бенчмарков), которые проверяют и оценивают способности языковых моделей в области общих знаний, логических рассуждений, здравого смысла и т. д. При этом актуальны исследование способностей и оценка качества языковой модели как системы объяснения в процессах поддержки принятия решений.

Рассмотрим основные модели объяснения, предложенные в философии науки.

1.1. Дедуктивно-номологическая модель объяснения

Согласно этой модели, научное объяснение состоит из двух основных элементов: экспланандума (*explanandum*), или объясняемого, – предложения, «описывающего явление, которое следует объяснить», и эксплананса (*explanans*), или объяснения, – «предложений, приводимых для объяснения явления» [32]. Чтобы объяснения (эксплананс) объясняли объясняемое (экспланандум), требуется выполнение нескольких условий.

• Объясняемое (экспланандум) должно быть логическим следствием объяснения (эксплананса), и предложения, из которых состоит эксплананс, должны быть истинными [32]; т. е. объяснение должно иметь форму дедуктивного аргумента, в котором экспланандум (объясняемое) является выводом из посылок, составляющих объяснение (эксплананс). В этом заключается «дедуктивный» компонент дедуктивно-номологической модели (ДН-модели).

• Эксплананс должен содержать по меньшей мере один «закон природы», который должен быть необходимой для вывода посылкой в том смысле, что при исчезновении этой посылки выведение экспланандума не будет валидным. Это номологический компонент модели (номологический – это философский термин, который, по сути, означает «законный») [32].

Формально постановка задачи объяснения для этой модели и для всех далее рассматриваемых моделей в самом общем виде выглядит следующим образом.

Модель объяснения – это кортеж <Λ, *Ex*1, *O*, *Ex*2>, где

• Λ – это множество законов; К. Гемпель [32] выделял несколько классов законов:

○ всеобщие законы, которые относятся ко всем областям знания L_c, – например, законы логики и математики;

○ частные законы отдельных наук L_s – например, законы химии, физики, биологии, социологии, политики и др.;

○ единичные факты L_f, они подчиняются более общим законам.

Таким образом, в ДН-модели множество законов Λ включает три определенных выше подмножества законов, т. е. $\Lambda = \{L_c, L_s, L_f\}$.

Понятие закона многозначно (закон природы, государственный закон и т. д.), поэтому приведем ряд определений, на которые в дальнейшем будем опираться в своих рассуждениях. Законом в философии называется необходимая связь (взаимосвязь, отношение) между событиями, явлениями, а также между внутренними состояниями объектов, определяющая их устойчивость, развитие, стагнацию или разрушение. В философском смысле под законом подразумевают «объективные связи явлений и событий, существующие независимо от того, известны они кому-нибудь или нет» [33].

Закон – это утверждение, выраженное словесно или математически, описывающее объективно существующие соотношения, связи между различными научными явлениями и объектами [28]. Закон предлагается в качестве объяснения фактов и признается на определенном этапе научным сообществом согласующимся с ними. Закон, справедливость которого была установлена не из теоретических соображений, а из опытных данных, называют эмпирическим законом.

• *Ex*1 – объясняемое, т. е. ситуация или явление которое необходимо объяснить.

• $O = \{o_j\}$ – множество фактов, которые характеризуют объясняемую ситуацию *Ex*1.

• *Ex*2 – объяснение.

Определение 1. ДН-объяснением некоторого явления или ситуации *Ex*1 будем называть отображение следующего вида: $DN{:}L_1(o_1)_{,\dots,}\, L_k(o_k) \to Ex2$, где $L_i()$ – закон, $L_i(.) \in \Lambda$; o_j – факты, они же переменные закона $L_i(.)$, все факты истинны, т. е. вероятность $Pr(o_j) = 1$; *DN* – процедура дедуктивного вывода из законов L_i и фактов o_j объясняемого явления *Ex*1. ♦

Объяснение, удовлетворяющее определению 1, называют каузально-номологическим объяснением. Это определение предполагает применение в объяснении детерминированного закона. Однако законы из областей квантовой физики, биологии (генетики), а также социологии и других гуманитарных наук имеют вероятностных характер. Поэтому в рамках ДН-объяснений определено дедуктивно-статистическое и индуктивно-статистическое объяснение.

Определение 2. Если в ДН-объяснении при определении объясняющих фактов в качестве закона выступает общий статистический закон $L_i^*(.) \in \Lambda$, такое объяснение называется дедуктивно-статистическим (*SN*-объяснением). ♦

В этом случае объясняемый объект или ситуация *Ex*1 может включать статистические характеристики случайного процесса (выборки и т. д.), математическое ожидание $M(o_j)$ и среднеквадратическое отклонение $\sigma(o_j)$, и должна быть определена номологическая вероятность объяснения $Ex2(M(o_j), \sigma(o_j))$.

В этом случае отображение объяснения из определения 1 запишем следующим образом:

$$SN{:}\ L_1(o_1),\dots, L_i^*(M(o_j), \sigma(o_j)),\dots,$$
$$L_k(o_k) \to Ex2(M(o_j), \sigma(o_j)), L_i^*(.) \in \Lambda.$$

Определение 3. Если в ДН-объяснении хотя бы один из фактов получен в результате индуктивного рассуждения (определена субъективная вероятность этого факта), то объяснение называется индуктивно-статистическим (*IN*-объяснением). ♦

Считается, что индуктивный факт должен иметь определенную субъективную вероятность. Обычно порог вероятности устанавливают на уровне 0,5, т. е. $o_j \mid Pr(o_j) > 0{,}5$.

В этом случае отображение объяснения из определения 1 запишем следующим образом:

$$IN{:}\ L_1(o_1),\dots, o_j \mid Pr(o_j) > 0{,}5\ ,\dots,$$
$$L_k(o_k) \to Pr(Ex1), L_i(.) \in \Lambda$$

В отличие от дедуктивно-статистических объяснений, которые формируются на основе большого количества наблюдений случайных величин и могут считаться реальным статистическим законом, индуктивные заключения формируются на

основе малого числа наблюдений и субъективны. В ДН-модели объяснения индуктивные факты не являются надежными без указания контекста их получения и оценки вероятности. Дедуктивно-статистические и индуктивно-статистические объяснения К. Гемпель [32] называет статистически релевантными и, соответственно, индуктивно релевантными.

В ДН-модели объяснения предложены критерии валидности объяснения. Первый критерий – это критерий обобщения, а второй – критерий симметричности объяснения. Оставаясь в рамках математических формулировок ДН-модели, определим обобщение элементов объяснения. В самом общем виде обобщение – это замена объясняемого понятия $Ex1$ и объясняющих фактов $(o_1, \ldots, o_k)$ именами классов или категорий, к которым они принадлежат. Обозначим имя класса объясняемого понятия $Kl(Ex1)$, а имена классов фактов через $Kl(o_j)$.

Определение 4. Обобщением ДН-модели объяснения будем называть объяснение, полученное путем замены имени понятия $Ex1$ на имя его класса $Kl(Ex1)$ и замены имен фактов $(o_1, \ldots, o_k)$ на имена их классов $(Kl(o_1), \ldots, Kl(o_k))$. ♦

Критерий валидности объяснения 1: ДН-объяснение DN: $L_1(o_1), \ldots, L_k(o_k) \rightarrow Ex2$ (см. определение 1) валидно (верно), если истинно его обобщенное ДН-объяснение DN: $Kl(L_1(o_1)), \ldots, Kl(L_k(o_k)) \rightarrow Kl(Ex2)$. ♦

Формально доказать валидность объяснения на основе обобщения можно, если закон, на основе которого строится объяснение, относится к общим законам L_c и имеется математическая модель объекта или ситуации. В этом случае может существовать строгое математическое или логическое доказательство ДН-объяснения, например, в логике предикатов и др. Однако во многих случаях, к примеру, в области гуманитарных наук, таких моделей просто не существует или они слишком абстрактны. Тогда этот критерий валидности объяснения может быть подтвержден или опровергнут экспертным способом, подстановкой в объяснение имен классов понятий объясняемого и фактов. Этот критерий позволяет утверждать о верности и всеобщности выбранного для объяснения закона.

Другой критерий валидности объяснения – это критерий симметричности логического вывода.

Критерий валидности объяснения 2: ДН-объяснение DN:$L_1(o_1),\ldots, L_k(o_k)\rightarrow Ex2$ (см. определение 1) валидно (верно), если истинно обратное ДН-объяснение, т. е. DN: $Ex2\rightarrow L_1(o_1),\ldots, L_k(o_k)$. ♦

Существование обратного вывода объяснения говорит о его непротиворечивости. Формально проверить этот критерий валидности можно в случаях, когда для объяснения используется известный всеобщий закон, известна модель объекта или ситуации. Для этого можно опираться на математический аппарат теории доказательств [34]. В других случаях возможна экспертная проверка непротиворечивости дедуктивного вывода, приведенного в объяснении.

Модель ДН-объяснения считается классической моделью научного объяснения объектов или ситуаций. Однако эта модель объяснения при условии соблюдения критериев валидности объяснения получается достаточно громоздкой и трудно обозримой. Полное и подробное объяснение, полученное с помощью ДН-модели, считается идеальным научным объяснением, которое называют «скрытой структурой» объяснения.

У философов существует точка зрения, согласно которой подробное объяснение избыточно, и поэтому считаются допустимыми так называемые «наброски объяснения» [32]. Допускается, что неполное объяснение, полученное, например, путем обобщения идеального объяснения, будет содержать объясняющий идеальное объяснение компонент. В данном случае речь идет об объяснении идеального объяснения в понятных терминах, например, «здравого смысла». Здравый смысл (*common sense*) – это совокупность навыков, форм мышления, взглядов на окружающую действительность, выработанных и используемых человеком в повседневной практической деятельности, которые разделяют окружающие люди.

Обычно объяснение считают чем-то, обеспечивающим понимание. В связи с этим одной из задач теории объяснения является выявление тех структурных особенностей объяснений, которые обеспечивают понимание. Понимание – это универсальная операция мышления, связанная с усвоением нового содержания, включением его в систему устоявшихся идей и представлений [35]. Считается, что наброски объяснения в терминах здравого смысла могут обеспечить понимание явления.

Будем считать здравым смыслом некоторые эмпирические законы, принятые обществом (социумом). Добавим в ранее рассмотренное в ДН-модели множество законов Λ законы здравого смысла L_{cs}, $\Lambda = \{L_c, L_s, L_f, L_{cs}\}$.

Тогда объяснение в рамках стратегии скрытой структуры представим следующим образом.

Определение 5. Объяснением $Ex2^*$ скрытой структуры ДН-объяснения $Ex2$ некоторого объекта или ситуации будем называть отображение следующего вида: DN^*: $L_{cs}^{\,1}(L_1(o_1)),\ldots, L_{cs}^{\,k}(L_k(o_k))\rightarrow Ex2^*$, где $L_{cs}^{\,j}()$ – законы здравого смысла $L_{cs}^{\,j}(.)\in\Lambda$; $Ex2^*$ –

объяснение здравого смысла; DN^* – процедура дедуктивного вывода из законов здравого смысла и фактов объясняемого явления. ♦

Объяснение скрытой структуры можно рассматривать как интерпретацию идеального объяснения ДН-модели. Проверить валидность объяснения скрытой структуры можно, если между основными законами L_c или частными законами L_s и законом здравого смысла L_{cs} определено отношение гомоморфизма, т. е. Ф: $L_c \rightarrow L_{cs}$ или Ф: $L_s \rightarrow L_{cs}$. При гомоморфизме элементы и отношения законов L_c, L_s отображаются в элементы и отношения законов здравого смысла L_{cs} при сохранении каузальных отношений. При гомоморфизме обратное отображение не однозначно.

При этом критерии валидности 1 и 2 могут не выполняться. Будем называть такое объяснение эмпирическим объяснением.

Философы считают, что объяснения в терминах законов здравого смысла (объяснения на подобии, аналогии, метафорах и т. д.) играет важную вспомогательную роль при формулировании идеального научного объяснения. Этот факт можно объяснить неоднозначностью обратного отображения Φ^{-1}: $L_{cs} \rightarrow \{L_{ci}\}$, т. е. из объяснения, основанного на законах здравого смысла (L_{cs}), можно получить множество строгих ДН-объяснений, основанных на общих законах $\{L_{ci}\}$. Это создает возможность выбора лучшей модели объяснения и активизирует мышление исследователя.

1.2. Унификационистская теория объяснения

Следующая модель объяснения основана на унификационистской теории, смысл которой заключается в том, что научное объяснение может быть представлено в виде единого описания различных объектов и явлений. В этой теории определено схематическое предложение следующего вида [36]: «Для всех X, если X есть O и A, то X есть P.» Здесь X – это переменная (объясняемое); O, A, P – это переменные, факты, например свойства объясняемого.

Левая часть этого предложения – посылка ($Prm = (\forall X(O, A))$), а правая – это вывод – объяснение ($Inf = (X \rightarrow P)$). Автор работы [36] вводит понятие схематических аргументов – это последовательности схематических предложений и классификатор, позволяющий определить, какие аргументы являются посылками, а какие выводами и правилами вывода.

Определено также и правило заполнения (замены) абстрактных переменных (X, O, A, P) предметными переменными (понятиями предметной области). Вводится определение паттерна аргумента, включающего аргументы, их классификацию и инструкцию по заполнению. Считается, что заполненный предметными переменными паттерн является объяснением. Конечно, можно получить множество паттернов объяснения, но какое из них верное?

Автор вводит понятие объяснительного резерва [36] – набора паттернов аргументов, включающего набор убеждений, разделяемых учеными в конкретный период времени. Для доказательства, что конкретный вывод – паттерн объяснения – является здравым или приемлемым объяснением, нужно будет показать, что он принадлежит к объяснительному резерву.

Процедура вывода в данном случае заключается в подстановке в паттерн аргумента значений предметной переменной и проверке заполненного паттерна на соответствие объяснительному резерву. Если такое соответствие есть, то объяснение получено. В противном случае паттерн заполняется новыми данными. Строго говоря, схематическое предложение этой модели является описанием вывода в логике первого порядка. В логике первого порядка в аргументе посылки явным образом присутствует некоторый закон, на основании которого осуществляется вывод. Например, в утверждении «все люди смертны» (это закон), «Сократ – человек» (это факт), «Сократ смертен» (это заключение – вывод) вывод получен на основе закона.

В унификационисткой модели объяснения явной ссылки на некоторый закон нет. Ссылка в объяснении на закон обеспечивает важное свойство объяснения – его каузальность. Это вызвало философские споры о корректности такой модели объяснения. Однако автор работы [36] приводит, во-первых, примеры объяснений, которые не содержат каузальных отношений, а во-вторых, он считает, что паттерны аргументов посылок и выводов должны быть обобщением исторического опыта человека и общества в различных сферах деятельности, отражать его культурные традиции. В результате длительного эволюционного процесса в социуме формируются убеждения, включающие в том числе и каузальные отношения. Все эти общественные убеждения включаются в объяснительный резерв предложенной модели и используются для валидации паттерна объяснения. В целом эта модель объяснения не лишена недостатков, но если, соглашаясь с автором этой модели, будем считать общественные убеждения некоторым эмпирическим социальным законом L_{so} и добавим их во множество законов Λ классической модели объяс-

нения, то получим значительное расширение модели объяснения, $\Lambda = \{L_c, L_s, L_f, L_{cs}, L_{so}\}$.

Определение 6. Эмпирическим объяснением некоторого объекта или ситуации $Ex1^{emp}$ будем называть отображение вида DN^{emp}: $L_{so}{}^{1}(o_1)$, ..., $L_{so}{}^{k}(o_k) \rightarrow Ex2^{emp}$, где $L_{so}{}^{i}()$ – эмпирический социальный закон $L_{so}{}^{i}(.) \in \Lambda$; o_j – факты, переменные закона $L_{so}{}^{i}(.)$, все факты истинны, вероятность $Pr(o_j) = 1$; $Ex2^{emp}$ – объяснение, основанное эмпирических законах; DN^{emp} – процедура подстановки эмпирического объяснения $Ex2^{emp}$ из множества объяснительного резерва ER, $Ex2^{emp} \in ER$. ♦

Критерии валидности 1 и 2 для эмпирического социального объяснения в данном случае можно применять экспертным способом. Однако из-за эмпиризма закона L_{so} нельзя ничего сказать (даже дать вероятностную оценку объяснения) о достоверности объяснения. В данном случае выполнение этих критериев будет свидетельствовать о корректности и непротиворечивости вывода объяснения, но не о его достоверности. В качестве объяснительного резерва могут выступать разнообразные справочники и энциклопедии.

1.3. Прагматические теории объяснения

Ранее были рассмотрены модели объяснения, которые не учитывали роль человека в процессе объяснения. Существуют прагматические теории объяснения, в которых человек непосредственно включен как в процесс генерации объяснения, так и в процесс потребителя объяснения – в качестве потребителя.

Теория объяснения содержит «прагматические» элементы, если эти элементы требуют обязательной отсылки к фактам об интересах, убеждениях или других психологических характеристиках тех, кто дает объяснение или его получает, и обязательна отсылка к контексту, в котором возникает объяснение [37].

Авторы классической ДН-модели объяснения согласны с тем, что прагматические элементы играют какую-то роль в процессе предоставления или получения объяснений, однако они полагают, что существует непрагматическое ядро объяснения, описание которого является главной задачей теории объяснения [32].

Современные прагматические теории объяснения основаны на конструктивном эмпиризме, изложенном в работе американского философа Б. ван Фраассена [37]. Он считает, что цель науки состоит в построении «эмпирически адекватных» теорий, дающих истинные или верные описания доступных наблюдению явлений, а научная деятельность является скорее конструированием, чем открытием: конструирование моделей, которые должны быть адекватны явлению, а не открытие истины, имеющей отношение к ненаблюдаемому [37]. «Цель науки – дать нам теории, которые являются эмпирически адекватными и принятие теории включает, как веру, только то, что она эмпирически адекватна» [37]. Под «эмпирической адекватностью» Б. ван Фраассен имел ввиду совпадение эмпирических проявлений теоретической модели явления и самого явления.

Теория конструктивного эмпиризма дополняет теорию научного реализма, основанную на строгих логических выводах, законах и доказательствах. Однако, как считает Б. ван Фраассен, строгость логических выводов в теории научного реализма основывается на недоказанных постулатах и аксиомах, что ограничивает ее применение в гуманитарных сферах. При построении теории конструктивного эмпиризма автор ссылается на модель семиотики – науки о знаках и знаковых системах, в которой наблюдаемое явление представлено на трех уровнях: синтаксическом, семантическом, прагматическом [38].

Строгое логическое объяснение, полученное в рамках теории научного реализма, можно представить как синтаксическую абстрактную структуру. В конструктивном эмпиризме осуществляется семантическая (смысловая) интерпретация таких логических структур.

В конструктивном эмпиризме человек – это наблюдатель проявлений действительности, конструктор эмпирической модели действительности и ее вадидатор. Следовательно, «принятие» теории означает всего лишь убеждение в ее эмпирической адекватности [37]. Вопросы применения конструктивного подхода для исследования социальных систем рассмотрены в работе [39].

В рамках теории конструктивного эмпиризма предложена модель объяснения. Эта модель объяснения представляется в виде тройки:

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

где Q – вопрос; T_k – тема или контекст; $X = \{X_1,..., X_n\}$ – контрастные классы ответов, по сути альтернативы ответов; RL – отношение релевантности темы и альтернативного ответа.

Поясним эту модель. Объяснение – это не просто набор суждений, а всегда ответ на вопрос «почему?», который всегда возникает в определенном контексте и определяется тремя факторами. Первым фактором является тема – то, о чем вопрошается (T_k). Вторым фактором будет выступать контрастный класс, представляющий собой класс, состоящий из множества альтернативных теме вы-

сказываний. Третий фактор – отношение релевантности между темой вопроса и контрастным классом, устанавливающее, что может выступить в качестве объяснения.

Пусть для конкретной темы T_k и заданного множества альтернативных ответов X может быть образовано множество пар – тема и альтернативный ответ – путем прямого произведения $T_k \times X$.

Определение 7. Для пары $(T_k , X_j) \in T_k \times X$, $X_j \in X$, существует отношение релевантности $RL \subseteq T_k \times X$, если субъект верит и убежден, что X_j отвечает на вопрос Q в контексте T_k. В этом случае объясняемое Q и объяснение X_j эмпирически адекватны. ♦

Высказывание X_j является релевантным вопросу Q тогда и только тогда, когда существует отношение RL для пары (T_k, X_j). Установление отношения релевантности не может быть задано однозначно. Например, рассмотрим вопрос Q = «Почему кровь циркулирует в теле?» и два возможных варианта ответа на него: X_1 = «потому что сокращение сердца заставляет кровь двигаться по артериям»; X_2 = «чтобы доставить кислород ко всем тканям организма».

Видно, что есть две альтернативы ответов (два контрастных класса), но вне контекста постановки вопроса нельзя установить отношение релевантности между темой вопроса Q и контрастным классом X [37].

Вопрос «почему?» фиксирует некую проблему и задает определенный контекст. При этом ответ на него включает и теоретический контекст – научное объяснение. Для данного контекста вопроса эмпирически адекватным будет ответ X_1 = «потому что сокращение сердца заставляет кровь двигаться по артериям». Здесь имеет место каузальное отношение релевантности. Однако если поменять контекст вопроса Q = «Зачем кровь циркулирует в теле?», то эмпирически адекватным будет ответ X_2 = «чтобы доставить кислород ко всем тканям организма», с функциональным отношением релевантности.

Считается, что «научное объяснение относится не к сфере чистой науки, а к сфере применения науки. Оно представляет собой использование науки для удовлетворения некоторых наших желаний; это всегда конкретные желания, возникающие в конкретном контексте, но они всегда являются желаниями получения определенной дескриптивной информации... Точное содержание желания и оценка того, насколько хорошо оно удовлетворено, разнятся от контекста к контексту» [37].

Обобщая рассмотренные модели объяснения, отметим основные отличия прагматических моделей объяснения от семейства ДН-моделей. Напомним, что семейство ДН-моделей объяснения (в литературе их называют классическими) разрабатывалось в рамках теории научного реализма, и такие модели характеризуются следующими свойствами:

– объяснение основывается на законах (детерминированных, статистических, эмпирических),

– логический вывод объясняемого через наблюдаемые факты основан на законах и должен быть корректным и истинным (достоверным),

– объяснение отражает объективные законы природы и (или) общества и не зависит от психологических особенностей субъекта (его интересов, убеждений, желаний, оценок и т. д).

Прагматические теории объяснения основаны на теории конструктивного эмпиризма и характеризуются следующими свойствами:

- субъект включен в модель объяснения. Объяснение формируется с учетом его психологических особенностей и контекста, который он формулирует в вопросе;
- ответ формируется на основе субъективной эмпирической адекватности объяснения. Утверждение об эмпирической адекватности объяснения гораздо слабее утверждения истинности (достоверности) объяснения в моделях ДН-объяснений. Однако эмпирическая адекватность объяснения позволяет удовлетворить исследовательские потребности субъекта;
- объяснение содержит научно обоснованные утверждения на уровне семантики – смысла объясняемого, а не логического вывода в терминах, доступных ограниченному кругу узких специалистов, как это делается в рамках теории научного реализма;
- объяснение направлено на удовлетворение потребностей, желаний конкретного субъекта в интересном ему контексте для получения дополнительной релевантной информации и расширения его мировоззрения – ментального пространства.

Эти два класса моделей образуют конкурирующие теории объяснения, но в процессах поддержки принятия решений могут дополнять друг друга.

Отметим, что в литературе приводятся модели объяснения для разных предметных областей, например модели объяснения в социологии, технике, математике, модели объяснения сознания и др. Выше были рассмотрены некоторые философские модели объяснения из методологии научного познания. Другие модели научного объяснения можно найти в философской литературе, например в обзорной работе [9]. Задача приведенного здесь краткого обзора заключается в том,

чтобы выделить наиболее известные, качественно отличающиеся классы моделей объяснения с тем, чтобы в дальнейшем можно было экспертным путем классифицировать тексты, сгенерированные большой языковой моделью, в один из классов научного объяснения. Это позволит оценить качество объяснения языковой модели и, следовательно, ее полезность для решения задач поддержки принятия решений.

1.4. Цель объяснения в поддержке принятия решений

Отметим, что, литература, посвященная различным конкурирующим моделям объяснения, достаточно обширна. Здесь исследователи поразному трактуют само понятие объяснения, однако достаточно мало уделяют внимания вопросу о целях объяснения – тому, для чего используются объяснения [31]. Очевидно, что цель объяснения определяется целью научного исследования, которое зависит от объекта исследования. Часто предлагаемыми в самом общем виде целями исследования считаются: доказательство, исследование структуры, принципа функционирования, прогноз развития, управление сложным объектом и др. Цель объяснения также зависит от области исследования; например, в практике обучения математике выделяют такие виды объяснения: объяснение – обоснование, объяснение решения задачи, объяснение – раскрытие смысла и объяснение – верификация.

Однако объединяющим началом разных моделей объяснения в разных предметных областях принято считать связь объяснения и понимания. Считается, что объяснение обеспечивает понимание. Под пониманием подразумевается универсальная операция мышления, связанная с усвоением нового содержания (новое содержание включено в объяснение), включением его в систему устоявшихся идей и представлений (по сути, в систему знаний субъекта) [35].

Фактически объяснение – это описание основных свойств исследуемого объекта или ситуации с возможной привязкой их к известным законам, которое запускает универсальную операцию мышления – понимание, обеспечивающее встраивание этой новой информации в уже имеющиеся у субъекта знания.

В работе [40] автор, связывая понимание и объяснение, считает, что тот, «кто понимает объяснение, постоянно и легко переходит от реального объекта к его «идеальной» модели (именно «идеальная» модель объекта создается абстракциями) и обратно», может повторить все познавательные операции объясняющего и осознать, почему произведены именно эти операции и в такой последовательности. Таким образом, объяснение обеспечивает связь реального мира с ментальными процессами мышления человека и направлено на формирование абстрактной идеализированной ментальной модели объекта или ситуации. Затем ментальная модель может быть выражена на строгом математическом языке и стать законом.

В этом случае в аспекте поддержки принятия решений можно несколько иначе сформулировать цель объяснения. Во-первых, объяснение направлено на формирование информационной среды об объекте или ситуации, включающей параметры объекта или явления, структуру, принадлежность к определенному классу (классификацию) и др. в ментальной сфере субъекта. В этом случае работают разнообразные отношения релевантности объяснения, отвечающие потребностям субъекта, и нестрогая эмпирическая адекватность объяснения.

Во-вторых, объяснение направлено на формирование или выявление каузальных отношений, строгий логический вывод на основе известных или вновь открытых законов или закономерностей для принятия решения по управлению объектом или ситуацией.

Таким образом, цель объяснения в процессах поддержки принятия решений имеет две составляющие:

– исследовательскую, направленную на формирование информационной среды для принятия решений и выработки альтернатив решений;

– практическую, направленную на принятие, обоснование и реализацию альтернатив решения по управлению объектом или ситуацией.

В этом случае конкурирующие модели объяснений – классическая (ДН-модель) и прагматическая дополняют друг друга. Если рассматривать процесс принятия решений как исследовательский процесс, то можно определить два этапа этого процесса. Первый этап – исследовательский – заключается в серии последовательных вопросов и ответов-объяснений, он позволяет сформировать необходимую информационную среду для выработки альтернатив принятия решений. На этом этапе уместны прагматические модели объяснения, не ограничивающие желания и потребности исследователя и позволяющие получить эмпирически адекватные объяснения альтернатив решения.

На втором этапе осуществляется выбор лучшей альтернативы с использованием ДН-модели объяснения, позволяющей обосновать альтернативы ссылками на известные законы и строгий логический вывод.

2. БОЛЬШИЕ ЯЗЫКОВЫЕ МОДЕЛИ И ОЦЕНКА ИХ СПОСОБНОСТЕЙ

Большие языковые модели представляют собой нейронные сети глубокого обучения, которые обучаются на гигантских объемах текстовых данных. В их основе лежит архитектура трансформера, которая включает набор нейронных сетей, состоящих из энкодера и декодера**.** Трансформер – это архитектура нейронной сети, направленная на решение задачи обработки естественного языка и многих других задач машинного обучения [41]. Энкодер – это элемент архитектуры трасформера, преобразующий корпус текста в векторные представления, в которых каждое слово представлено в виде вектора вероятностей его совместного употребления с другими словами корпуса текста, при этом сохраняется информация о структуре и взаимосвязях слов. Декодер использует закодированную информацию для генерации ответа или выполнения предсказаний. Он «раскодирует» полученные данные, создавая новые последовательности текста, учитывающие предыдущие слова, контекст и вероятности совместного употребления слов в этом контексте.

Нейронная сеть большой языковой модели представляется в виде многослойной структуры, в которой каждый слой состоит из множества искусственных нейронов, связанных с нейронами соседних слоев. Число слоев в разных больших языковых моделях отличается и может достигать нескольких десятков или больше. В результате обучения нейронной сети большим корпусом текста формируется словарь токенов – отдельных букв или их групп, слов, фраз, предложений, для которых определена вероятность их совместного употребления в определенном контексте. При кодировании входной информации и передачи ее от слоя к слою используется механизм внимания, позволяющий выявить важные элементы структуры языка. Механизмы внимания могут выделять различные типы синтаксических отношений между словами, фактически выделяя синтаксическую структуру предложения [42]. В результате обучения образуется векторное пространство контекстных векторов токенов, для которых определена (косинусная) мера близости токенов. При декодировании для ответа на вопрос из векторного пространства выбираются токены (слова, фразы, предложения), близкие к теме вопроса, из которых выстраиваются предложения – ответ на вопрос. Выбор из векторного пространства токенов, близких к токенам вопроса, позволяет сконструировать ответ на поставленный вопрос из слов, часто встречающихся в контексте вопроса.

Таким образом, большая языковая модель формирует статистический ответ на поставленный вопрос, не работая на уровне семантики, не анализируя смысл вопроса и ответа. Смысл ответа зависит от качества корпуса текста. Если в корпус включались тексты из проверенных источников, то ответ будет по смыслу верен и понятен.

Если языковая модель не понимает вопроса или была обучена неполными или ошибочными данными, то она пытается угадать ответ, опираясь на имеющиеся синтаксические шаблоны, что может привести к ложным ответам. Такое поведение большой языковой модели называется галлюцинацией. Галлюцинация – это феномен больших языковых моделей, когда они генерируют фактически неверные или вымышленные данные, которые не основаны на реальной информации. Такое поведение свойственно и людям, когда в условиях дефицита информации или времени для изучения проблемы, они формируют случайные или вымышленные ответы, далекие от реальности.

Проблема галлюцинаций больших языковых моделей особенно актуальна в приложениях, в которых достоверная информация критически важна. В настоящее время разработаны различные метрики оценки склонности языковых моделей к галлюцинациям и методы уменьшения галлюцинаций языковых моделей, связанные: с созданием корректных корпусов текста (ручная проверка, чистка текста); формированием корректных запросов к языковой модели, подсказкой правильного ответа; дообучением языковой модели; поиском верной релевантной информации во внешних источниках информации.

Необходимо отметить, что галлюцинации языковой модели в случаях решения исследовательских и творческих задач могут стимулировать интуицию человека и привести к оригинальному решению. Напомним, что эвристические методы решения творческих задач, такие как мозговой штурм, синектика, не отвергают абсурдные и контрпродуктивные альтернативы решений, считая их стимуляторами интуиции, способной привести к новому и оригинальному решению.

Несмотря на этот недостаток, языковые модели находят применение благодаря их способностям решать многие задачи, представленные на естественном языке, которые ранее решал только человек.

Большие языковые модели оцениваются с помощью специальных программ – бенчмарков, которые измеряют их основные качества: объем знаний, точность ответов, надежность и др.

Бенчмарк использует определенные наборы данных, метрики и задачи оценки для тестирования языковой модели, что позволяет сравнивать различные модели и измерять их точность. Выделяют бенчмарки на проверку знаний, логического мышления, понимания прочитанного текста, здравого смысла и др. [43]. Разработано большое количество бенчмарков; рассмотрим некоторые из них.

Бенчмарки на проверку знаний тестируют модели в различных областях. Они оценивают, насколько эффективно модель может вспоминать информацию из разных сфер, таких как физика, география и т. д. Известный бенчмарк MMLU (*Measuring Massive Multitask Language Understanding*) создан для проверки уровня фактических знаний модели по различным темам, таким как гуманитарные науки, социальные науки, история, компьютерные науки и даже право. Он включает 57 вопросов и 15 тыс. задач, направленных на то, чтобы убедиться в высоких способностях языковой модели. На этом бенчмарке языковая модель GPT-4-omni правильно ответила на 88,7 % заданных ей вопросов.

Бенчмарки на проверку логического мышления тестируют способности модели «думать» пошагово и делать логические выводы.

Используются бенчмарки оценки математических способностей языковых моделей. Например, тест GSM8K состоит из 8,5 тыс. задач по математике для средней школы. Решение этих задач требует выполнения моделью нескольких шагов последовательности элементарных вычислений. Языковые модели, специально обученные для математического рассуждения, показывают хорошие результаты на этом бенчмарке – например, модели GPT-4 достигают 96,5 % точности.

Бенчмарк вопросов и ответов (англ. *Graduate-Level Google-Proof Q&A Benchmark,* GPQA) уровня аспирантуры оценивает логическое мышление языковой модели, используя набор данных всего из 448 вопросов. Этот сложный тест, разработанный экспертами из областей биологии, физики и химии, языковая модель GPT-4-omni проходит, набирая лишь 53,6 % точности, в то время как аспиранты достигают 65 %.

Исследования по разработке рассуждающих языковых моделей ведутся многими основными конкурирующими производителями языковых моделей. Так, в 2024 г. компания OpenAI выпустила новую языковую модели OpenAI-o1, которая демонстрирует отличные результаты в сложных рассуждениях, превосходя людей в тестах по математике, кодированию и естественным наукам. На отборочных экзаменах Международной математической олимпиады (англ. *International Mathematical Olympiad,* IMO) эта модель правильно решила 83 % задач, тогда как ее предшественница GPT-4o дала лишь 13 % верных ответов. Разработчики из OpenAI утверждают, что при решении сложных контрольных задач по физике, химии и биологии модель демонстрирует результаты, сопоставимые с результатами аспирантов.

Бенчмарки на понимание прочитанного текста тестируют способности модели интерпретировать естественный язык и генерировать соответствующие ответы. Тестирование заключается в получении ответов на вопросы по текстам, что позволяет оценить понимание, способность делать выводы и улавливать и не забывать важные детали.

Одним из бенчмарков для тестирования понимания прочитанного является тест DROP (*Discrete Reasoning Over Paragraphs*), который ставит перед моделями задачу выполнения рассуждений на основе анализа содержимого абзацев. Этот бенчмарк включает 96 тыс. вопросов для проверки способностей языковой модели к рассуждению. Вопросы DROP содержат информацию, которая требует от моделей выполнения математических операций, таких как сложение, вычитание и сравнение, на основе информации, разбросанной по всему тексту. Отвечая на эти сложные вопросы, языковая модель GPT-4 достигает точности 80 %, в то время как люди дают 96 % правильных ответов на наборе данных DROP.

Бенчмарки на проверку общих знаний (здравого смысла) оценивают способность модели к обобщенным знаниям о мире. Такие наборы тестов обычно включают вопросы, требующие для правильного ответа обширных энциклопедических знаний. Тестирование здравого смысла в языковых моделях оценивает способность модели делать суждения и выводы, соответствующие человеческому мышлению. Люди формируют целостное представление о мире через практический опыт, а языковые модели обучаются на огромных наборах данных, не понимая контекста.

Бенчмарк HellaSwag (*Harder Endings, Longer Contexts, and Low-Shot Activities for Situations with Adversarial Generations*) предназначен для проверки способности модели предсказывать правдоподобное продолжение определенного сценария. Те-

стирования методами HellaSwag показывают, что современные модели, такие как GPT-4, достигли уровня точности, близкого к человеческой.

Бенчмарк IFEval (*Inference and Fidelity Evaluation*) позволяет оценить как точность, так и качество сгенерированного текста. Сначала модель оценивается по показателю Inference – способности генерировать текст на большом объеме данных. Затем происходит оценка качества (Fidelity) сгенерированного текста. Оценка включает в себя проверку на соответствие сгенерированного текста ожидаемому результату, а также оценку степени сохранения смысла и структуры текста. Далее вычисляется итоговая оценка IFEval, которая учитывается как способность модели генерировать текст, так и качество этого текста. Чем выше оценка IFEval, тем лучше модель справляется с задачей генерации текста.

Бенчмарк HumanEval – это эталонный набор данных, предназначенный для объективной оценки качества кода, генерируемого моделями искусственного интеллекта на основании текстового описания задачи. Бенчмарк состоит из 164 задач по программированию, написанных вручную специально для этого набора данных, чтобы гарантировать их отсутствие в обучающих выборках моделей. Все задачи сформулированы на языке Python и представлены в виде фрагментов кода с описанием.

В табл. 1 представлены оценки производительности больших языковых моделей от ведущих разработчиков: GigaChat 2 MAX от российской компании Сбер; Qwen 2.5 72B от ведущей платформы электронной коммерции Alibaba; Llama 3.3 70B (разработчик – компания Meta AI); GPT-4o (разработчик – OpenAI); DeepSeek-V3 от китайской компании DeepSeek, принадлежащей фонду High-Flyer; Yandex GPT5 Lite Instruct от российской компании Яндекс. Тестирование проводилось на бенчмарках следующих категорий: общие знания, математика, работа с кодом, качество генерации текста. Данные актуальны [44, 45] на начало 2025 г. Цифры в таблице показывают процент решенных задач.

Тестирование способностей языковых моделей с помощью бенчмарков позволяет объективно оценить и сравнить качество моделей разных производителей, понять текущий уровень и динамику развития языковых моделей. Производители языковых моделей считают, что способности этих моделей приближаются к способностям человека, что позволяет им быть ассистентом в системах поддержки принятия решений.

Однако решение практических задач, например, по принятию решений в сложных экономических, политических или социальных ситуациях требует от ЛПР комплексного сочетания всевозможных способностей человеческого интеллекта. Поэтому автоматизированные метрики сами по себе не могут охватить весь спектр оценки языковой модели, особенно когда дело касается субъективных аспектов понимания и генерации языка. Здесь человеческая оценка является гораздо более точной.

В этом случае целесообразно для оценки привлекать экспертов или группу экспертов, способных дать, например, на основе опроса, точную и надежную оценку способностей больших языковых моделей [43].

В случае экспертного оценивания объяснений языковых моделей возникают некоторые проблемы. Дело в том, что необходимо оценить взаимодействие человека и большой языковой модели. Языковая модель выдает статистически правдоподобное объяснение, не понимая его смысла, а человек пытается понять объяснение и встроить его в собственную систему знаний. Понимание человеком объяснения языковой модели зависит от уровня его знаний, и поэтому разные люди, получая один и тот же ответ языковой модели, могут дать ему разные оценки.

В этом случае возникает задача разработки метода оценки индивидуальной (не групповой) удовлетворенности человека объяснениями ситуации

Таблица 1

Производительность больших языковых моделей от ведущих разработчиков

Категория	Название бенчмарка	GigaChat 2 MAX	Qwen 2.5 72B	Llama 3.3 70B	GPT-4o	DeepSeek-V3	Yandex GPT5 Lite Instruct
Общие знания	MMLU (RU)	80,46	78,30	65,08	80,00	73,74	70,0
	MMLU (EN)	86,00	83,85	78,57	88,70	85,24	75,8
Математика	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
Работа с кодом	HumanEval	87,20	86,60	86,00	84,00	91,46	71,8
Качество генерации текста	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

принятия решения, которые он получает от большой языковой модели. Решение данной задачи будет рассмотрено во второй части этой статьи.

3. МОДЕЛИ ОБЪЯСНЕНИЯ И БОЛЬШИЕ ЯЗЫКОВЫЕ МОДЕЛИ

Рассмотренные ранее модели объяснения образуют два класса моделей: классические (дедуктивно-номологические), основанные на строгом логическом выводе и известных объективных законах или закономерностях, и прагматические, учитывающие цели и желания исследователя или объясняющего, а также контекст вопроса, требующего объяснения.

В этих моделях сам человек формулирует объяснение, основываясь на наблюдаемых фактах и знаниях. Очевидно, что знания человека-исследователя, необходимые для формулировки ответа-объяснения, в каждом классе моделей объяснения будут разными. Бенчмарки показывают хорошие способности языковых моделей в разных областях. Способна ли языковая модель дать объяснение, которое удовлетворило бы ЛПР?

Знания языковой модели определяются содержанием корпуса текста, которым ее обучили. Тогда для работы языковой модели в классе классических моделей объяснения ее необходимо обучить известными законами логики, математики, физики, химии, биологии и т. д. В этом случае модель с большой вероятность будет генерировать строгие логические объяснения, обоснованные известными законами. Достоверность таких выводов может быть проверена, а сами выводы и объяснения можно использовать для управления ситуацией.

Для работы языковой модели, функционирующей в классе прагматических моделей объяснения, ее нужно обучать общими знаниями, законами здравого смысла, традициями и др. Тогда будут получаться эмпирически адекватные объяснения, основанные, согласно теории конструктивного эмпиризма, на вере исследователя. Такие объяснения полезны для расширения ментального пространства ЛПР, повышения его осведомленности и стимулирования интуиции для генерации нетривиальных творческих решений.

Обучение языковых моделей для получения классических объяснений или прагматических объяснений с «чистого листа» – это очень затратные процедуры. Во-первых, это сбор, подготовка, очистка обучающего корпуса релевантного задаче текста очень больших размеров. Подготовку корпусов текстов, как правило, выполняют специальные коллективы, нанятые по аутсорсингу. В словарях современных языковых моделей количество токенов исчисляется миллиардами. Считается, что только у очень больших языковых моделей появляются эмерджентные свойства, такие как, например, способность к рассуждению. Во-вторых, для обучения нейронной сети требуются очень большие вычислительные мощности – суперкомпьютеры – и, соответственно, большие временны́е и энергетические затраты. Обучение языковой модели с «чистого листа» доступно крупным компаниям, которые позволяют пользоваться уже обученными моделями другим разработчикам для решения их собственных задач.

Уже обученные языковые модели, имеющиеся в открытом доступе, можно дообучить собственными текстами, ориентируя их работу на решение собственных конкретных задач. В этом случае этапы подготовки корпуса текста и обучение модели также присутствуют, но уже с меньшими затратами.

Большая языковая модель – это огромная нейронная сеть с огромным количеством связей между токенами. Каждый вопрос к сети активирует определенную цепочку токенов, формируя ответ. Вариантов активизации цепочек токенов тоже огромное количество, а, следовательно, количество правдоподобных альтернативных ответов также велико.

В настоящее время активно развивается альтернативный способ настройки большой языковой модели под собственные задачи с помощью так называемых промптов (англ. *Prompt Engineering*, PE) [46, 47]. Промпт – это подсказка или последовательность подсказок языковой модели. Промпты представляются в виде алгоритма или сценария решения пользовательской задачи, создаются пользователем и могут встраиваться в языковую модель. Таким образом, промпты управляют работой языковой модели для получения лучшего результата.

Техники промптов и модели объяснения

Рассмотрим несколько наиболее популярных техник создания промптов.

Zero-shot. Это техника создания промпта без использования примеров входных или выходных данных, используемых для дообучения языковой модели. Это обычный простой вопрос языковой модели, в ответ на который она может сообщить информацию, для получения которой не требуется многошаговая инструкция.

Few-shot. Это техника создания промпта с примерами, т. е. используются несколько приме-

ров входных и выходных данных, чтобы дообучить модель. Языковая модель, обучаясь на этих примерах, выдает ответы, следуя их шаблонам. Например, так можно дообучить модель для точного определения тональности (эмоциональной «окраски») анализируемого текста. Few-shot-промптинг может использоваться в качестве техники для обеспечения контекстного дообучения языковой модели. В примерах задается контекст, в котором рассчитывают получить ответ.

Role based. Это техника создания промпта, в основе которой лежит задание роли языковой модели, в которой она будет выступать при генерации ответа на поставленный вопрос. Например, можно попросить модель ответить на некоторый вопрос с точки зрения экономиста, финансиста, юриста и т. д. Вопрос один и тот же, а ответы будут сгенерированы в контекстах разных предметных областей [48].

Chain-Of-Thought (CoT). Это техника создания промпта, который заставляет языковую модель думать поэтапно, шаг за шагом. В данной технике пользователь задает языковой модели алгоритм решения задачи. Это делает размышления модели близкими к человеческим. Как правило, сложная задача декомпозируется на подзадачи, последовательное решение которых повышает точность общего решения. Применение этой техники в больших языковых моделях существенно повышает точность решения математических и логических задач известного математического бенчмарка GSM8K.

Авторы этой техники констатируют повышение качества решения лингвистических задач и задач общих знаний и здравого смысла в случае применения этого промпта. Эта популярная техника в разных модификациях широко используется в рассуждающих языковых моделях [49].

Chain-of-Verification. Это техника создания промпта, которая дополняет Chain-Of-Thought, заставляя модель проверять все предыдущие шаги перед тем, как сделать следующий шаг, делая рассуждающие языковые модели более надежными [50].

Chain-of-Note. Это техника создания промпта, который заставляет языковую модель делать так называемые «заметки» в процессе решения задачи, поясняя каждый последовательный шаг. Эта техника позволяет обнаружить галлюцинации языковой модели [51].

Chain-of-Knowledge. Это техника создания промпта, в который включаются проверенные знания о предметной области, что позволяет языковой модели использовать их для решения задачи. В отличие от техники Chain-Of-Thought-промптинга, в этой технике языковая модель в ходе рассуждения опирается на известные факты, законы или закономерности, которые пользователь включает в промпт, и выстраивает из них логическую цепочку, приводящую к конкретному и обоснованному ответу. Например, для решения задачи по физике пользователь включает в промпт законы физики, на которые языковая модель будет опираться при генерации решения задачи. Корректность ответа будет выше, чем в простом Chain-Of-Thought [52].

Tree of Thoughts (ToT) [53]. Эта техника применяется для решения сложных исследовательских задач или задач стратегического планирования, когда традиционные или простые методы создания промптов оказываются недостаточными. В работе [53] предложен промпт Tree of Thoughts, который обобщает метод цепочки мыслей (CoT). Эта техника предлагает обращать внимание на исследование ответов, которые служат промежуточными шагами для общего решения проблем с помощью языковых моделей. Метод ToT поддерживает построение дерева ответов (дерева мыслей), в узлах которого расположены промежуточные тексты ответов. На построенном дереве предлагается применять алгоритмы поиска, например, поиска в ширину и поиска в глубину, для получения решения и его анализа.

ReAct Prompting. В работе [54] предложена техника ReAct, в которой для генерации цепочек рассуждений использовалась языковая модель в интерактивном режиме с человеком с целью формирования релевантных задаче действий. Генерация цепочек рассуждений позволяет модели создавать, отслеживать и обновлять планы действий, а также обрабатывать ошибочные ситуации. Действия позволяют взаимодействовать с внешними источниками информации, такими как базы знаний.

Промпт ReAct позволяет языковой модели взаимодействовать с внешними источниками для получения дополнительной информации, что приводит к более надежным и достоверным ответам. ReAct улучшает интерпретируемость и надежность языковой модели. В целом авторы обнаружили, что наилучшим подходом является использование ReAct в сочетании с цепочкой мыслей (CoT), что позволяет использовать как внутренние знания, так и внешнюю информацию, полученную в процессе рассуждения.

Промпты здравого смысла (Commonsense). В настоящее время большой исследовательский интерес представляют промпты для поддержки дедуктивных рассуждений на основе знаний здраво-

го смысла. Дело в том, что большие языковые модели обучаются на больших текстовых корпусах, а отвечая на вопросы, дают статистически лучшие ответы. Такие ответы верные, но они, как правило, не содержат элементы дедуктивного вывода, которые для человека кажутся очевидными и представляются в виде неявных знаний. Для получения от языковой модели ответа, содержащего объяснение с дедуктивным выводом, разрабатываются промпты, подсказки языковой модели для извлечения осмысленного ответа. Генерация такой подсказки языковой модели основана на использовании внешних баз знаний здравого смысла. Это статические базы знаний ConceptNet [55], ATOMIC [56] и динамическая база знаний здравого смысла COMeT [54], в которой знания здравого смысла формируются из контекста вопроса к языковой модели и представляются в виде графа рассуждений. Эксперименты с этой динамической базой знаний [57] показали рост производительности языковых моделей на бенчмарках здравого смысла. В работе [58] представлен метод генерации знаний из самой языковой модели и предоставления этих знаний в качестве дополнительных входных данных (подсказки) при ответе на вопрос. Этот метод не требует доступа к структурированной базе знаний здравого смысла, но при этом повышает производительность современных больших языковых моделей. В работе [59] рассмотрены вопросы генерации языковой моделью ответов здравого смысла на основе внутреннего диалога с языковой моделью. В этом методе не используются внешние базы знаний, но оценки производительности языковой модели на бенчмарках здравого смысла с таким промптом показали рост. Однако, как считают авторы работы [59], недостаток больших языковых моделей – это отсутствие у них интроспекции, т. е. знаний о своих знаниях.

Выше были рассмотрены лишь некоторые техники промптов. В настоящее время множество разработчиков предлагают пользовательские промпты, которые могут встраиваться в уже обученные языковые модели и не требуют дополнительного дообучения языковой модели на предметной области. Это представляется недорогим вариантом подстройки языковой модели для решения пользовательских задач.

В поддержке принятия решений существуют эвристические приемы (методы) анализа ситуаций в условиях неопределенности, которые представляются как методики поиска решений в сложных ситуациях. Это, например, метод «пять почему», направленный на выявление причин проблемы; метод «Фишбоун», позволяющий построить иерархическую модель причинно-следственных связей; метод гирлянд и ассоциаций, направленный на генерацию творческих решений; SWOT-анализ, позволяющий определить стратегию развития организации, и др. Все эти методики прошли апробацию в практике принятия решений, а алгоритмы их работы можно рассматривать как алгоритм работы промпта большой языковой модели.

Промпт можно представить как управляющую программу, которая позволяет извлечь в результате выполнения последовательных шагов из языковой модели необходимую для принятия решений информацию.

Рассмотрим, как перечисленные выше техники промптов могут помочь языковой модели в реализации рассмотренных моделей объяснения.

Для прагматических теорий объяснения, ориентированных на удовлетворение желаний и интересов пользователя, подойдут техники, позволяющие организовать диалог с языковой моделью для расширения ментального пространства и осведомленности пользователя. Напомним, что в прагматических моделях объяснения важно, чтобы между темой вопроса и объяснением языковой модели существовало отношение релевантности и убеждение ЛПР, что объяснение эмпирически адекватно.

Для поддержки прагматических моделей объяснения можно применить техники промптов, приведенные в табл. 2.

Отметим, что техника промпта Few-shot позволяет реализовать унификационистскую модель объяснения [36]. Напомним, что в этой модели задается шаблон, в который подставляется объяснение, выбираемое из объяснительного ресурса объяснения. Объяснительный ресурс – это мнение видных ученых, обычаи, традиции, сведения, которые могут быть представлены в справочниках, энциклопедиях и др.

Для классических теорий объяснения (ДН-моделей [32]**)**, направленных на получение и обоснование объяснения в виде логического вывода решения, основанного на известных законах и закономерностях, подойдут техники промпта, представленные в табл. 3.

Ранее было сказано о двух целях объяснения в поддержке принятии решений. Это исследовательская цель, которая достигается использованием языковой модели с промптами, реализующими получение прагматических объяснений, повышающих осведомленность эксперта, и практическая цель, которая достигается с использованием языковой модели с промптами для классических теорий объяснения, помогающими получить и обосновать строгий логический вывод решения и его применение для управления ситуацией.

Таблица 2

Роль промпта в прагматической модели объяснения

Техника промпта	Роль промпта в прагматической модели объяснения
Zero-shot	Позволяет получить ответ на любой вопрос
Few-shot	Позволяет получить ответ в виде заданного пользователем шаблона
Role based	Позволяет получить ответы на один и тот же вопрос в разных контекстах
Chain-Of-Thought (CoT)	Позволяет представить сложную задачу в виде связанных подзадач и решить их последовательно
Chain-of-Verification	Расширяет технику цепочки мыслей (CoT), позволяет верифицировать рассуждения языковой модели и обнаружить галлюцинации
Chain-of-Note	Расширяет технику цепочки мыслей (CoT), позволяет повысить качество рассуждений языковой модели, дополняя их пояснениями
Tree of Thoughts (ToT)	Позволяет построить дерево рассуждений вместо цепочки мыслей в методе (CoT); это позволяет найти лучшее альтернативное объяснение
ReAct Prompting	Допускает интерактивный режим пользователя и языковой модели. Пользователь может задействовать внешние данные для решения исследовательских задач и задач планирования

Таблица 3

Роль промпта в классической модели объяснения

Техника промпта	Роль промпта в классической модели объяснения
Chain-of-Knowledge	Позволяет дообучить языковую модель моделью знаний, включающую описание законов или закономерностей конкретной предметной области. Это позволяет повысить достоверность логического вывода языковой модели в исследуемой предметной области
ReAct Prompting	Организует интерактивный режим работы пользователя и языковой модели с возможностью обращения к внешним источникам проверенных данных. Это позволяет использовать описание законов из внешних источников данных, что может повысить достоверность логического вывода
Commonsense	Позволяет получить «набросок» классического объяснения в терминах здравого смысла. Такое объяснение выше было названо объяснением скрытой структуры дедуктивно-номологического объяснения

Применение комбинаций существующих техник промптов или разработка собственного промпта позволяют настроить большую языковую модель на решение задач объяснения альтернатив решений как в парадигме классической теории объяснений, так и в прагматической теории объяснений.

Вопросы разработки и тестирования пользовательских промптов для решения исследовательской и практической задач поддержки принятия решений в настоящей работе не рассматриваются.

ЗАКЛЮЧЕНИЕ

Рассмотрены вопросы применения больших языковых моделей для генерации и объяснения альтернатив решений, полученных системой поддержки принятия решений в условиях неопределенности. Рассмотрены классические (дедуктивно-номологические) и прагматические модели объяснения, предложенные философами. Сформулированы цели и задачи объяснения в процессах поддержки принятия решений в условиях неопределенности. Цель объяснения в системе поддержки принятия решений заключается в формировании информационной среды принятия решений и сводится к решению исследовательской задачи, позволяющей сформулировать альтернативы решений, и практической задачи, направленной на обоснование и реализацию лучшей альтернативы. Приведены концептуальный анализ функционирования больших языковых моделей, оценка их способностей при решении типовых тестовых задач, определяющие текущий уровень их способностей. Рассмотрены основные техники промптинга (системы запросов к языковой модели) позволяющие настроить языковую модель на генерацию объяснений для решения исследовательской и практической задач поддержки принятия решений в условиях неопределенности.

В первой части статьи сформулированы основные понятия и определения, которые далее будут использованы во второй части статьи. Вторая часть статьи посвящена вопросам измерения и оценки удовлетворенности лица, принимающего решения, объяснениями больших языковых моделей. Оценки удовлетворенности проводятся для исследовательской и практической задач объясне-

ния в условиях неопределенности. Будут проанализированы объяснения двух российских больших языковых моделей, которые будут отнесены к выделенным классам моделей объяснения.

ЛИТЕРАТУРА

1. *Simon, H.A.* Rationality as Process and as Product of Thought / In: Decision Making: Descriptive, Normative, and Prescriptive Interactions, ed. by D.E. Bell, H. Raiffa, A. Tversky. – Cambridge: Cambridge University Press, 1988. – P. 58–77.
2. *Checkland, P.B.* Systems Thinking, Systems Practice. – New York: Wiley, 1981. – 330 p.
3. *Axelrod, R.* The Structure of Decision: Cognitive Maps of Political Elites. – Princeton: Princeton University Press, 1976.
4. *Kosko, B.* Fuzzy Thinking: The New Science of Fuzzy Logic. New York: Hyperion, 1993. – 336 p.
5. *Кулинич А.А.* Верификация когнитивных карт на основе объяснения прогнозов // Управление большими системами. – 2010. – Вып. 30.1. – С. 453–469. [*Kulinich, A.A.* Cognitive Maps Verification Based on Processes Explanation // Large-Scale System Control. – 2010. – No. 30.1. – P. 453–469. (In Russian)]
6. *GPT-2* нейросеть от OpenAI. Быстрый старт. – URL: https://habr.com/ru/articles/440564/ (дата обращения: 16.02.2026). [*GPT-2* nejroset' ot OpenAI. Bystryj 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. – P. 275–276. – DOI:10.1177/14780771241280148
10. *Devlin, J., Chang, M.-W., Lee, K., 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/ (дата обращения: 14.02.2026). [*GigaChat*-2. – URL: https://giga.chat/ (Accessed February 14, 2026). (In Russian)]
15. *Yandex GPT* 5.1. – URL: https://ya.ru/ai/gpt (дата обращения: 14.02.2026). [*Yandex GPT 5*.1. – URL: https://ya.ru/ai/gpt (accessed February 14, 2026). (In Russian)]
16. *Визильтер Ю.В.* Приоритетные направления исследований и ключевые тенденции развития технологий ИИ // Труды Двадцать второй Национальной конференции по искусственному интеллекту с международным участием, КИИ–2025. – Т. 1. – Санкт-Петербург, 2025 г. – С. 7–25. – DOI: 10.15622/rcai.2025.001 [*Vizil'ter, Yu.V.* Prioritetnye napravleniya issledovanii i klyuchevye tendentsii razvitiya tekhnologii II // Trudy Dvadtsat' vtoroy Natsional'njy konferentsii po iskusstvennomu intellektu s mezhdunarodnym uchastiem, KII–2025. – Vol. 1. – St. Petersburg, 2025. – S. 7–25. – DOI: 10.15622/rcai.2025.001] (In Russian)
17. *Брагин А.В., Бахтизин А.Р., Макаров В.Л.* Большие языковые модели четвертого поколения как новый инструмент в научной работе // Искусственные общества. – 2023. – Т. 18, № 1. – DOI: 10.18254/S207751800025046-9 [*Bragin, A., Bakhtizin, A., 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. *Бахтизин А.Р.* Вопросы прогнозирования в современных условиях // Экономическое возрождение России. – 2023. – № 2 (76). – С. 53–62. [*Bakhtizin, A.R.* The Challenges of Forecasting under Current Conditions // The Economic Revival of Russia. – 2023. – No. 2 (76). – P. 53–62.] (In Russian)]
19. «*Коммерсантъ*» узнал о тестировании российских языковых моделей для «Госуслуг» // Forbes. – 02.02.2024. – URL: https://www.forbes.ru/tekhnologii/505447-kommersant-uznal-o-testirovanii-rossijskih-azykovyh-modelej-dla-gosuslug?ysclid=lzfs8wut7z80399139 (дата обращения: 14.02.2026). [«*Kommersant*» uznal o testirovanii rossijskih yazykovyh modelej dlya «Gosuslug» // 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. *Гусев А.* Обзор Российских систем искусственного интеллекта для здравоохранения. – URL: https://webiomed.ru/blog/obzor-rossiiskikh-sistem-iskusstvennogo-intellekta-dlia-zdravookhraneniia/ (дата обращения: 14.02.2026). [*Gusev, A.* Obzor Rossijskih sistem iskusstvennogo intellekta dlya zdravoohraneniya. – 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., von Krogh, G.* Organizational Decision Making Structures in the Age of Artificial Intelligence // California Management Review. – 2019. – Vol. 61, no. 4. – P. 66–83.
22. *Gasanov, E.* Decision Intelligence: искусственный интеллект с человеческим лицом // IT World. – 22.02.2022. – URL: https://www.it –world.ru/cionews/1q746pou69z40kokwskckk4gkg4c0og.html (дата обращения: 14.02.2026). [*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. *Виртуальные* помощники // TAdviser. – 06.08.2022. – URL: https://www.tadviser.ru/index.php (дата обращения 15.02.2026) [*Virtual'nye* pomoshchniki // TAdviser. – August 6, 2022. – https://www.tadviser.ru/index.php (accessed February 15, 2026). (In Russian)]
24. *ИИ* в аналитике: что за пределами BI? // TAdviser. – 04.08.2022. – URL: https://www.tadviser.ru/index.php (дата обращения: 15.02.2026). [*II* v analitike: chto za predelami BI? // TAdviser. – August 4, 2022. – URL: https://www.tadviser.ru/index.php (accessed February 15, 2026). (In Russian)]

25.*Дудихин В.В., Кондрашов П.Е.* Методология использования больших языковых моделей для решения задач государственного и муниципального управления по интеллектуальному реферированию и автоматическому формированию текстового контента // Государственное управление. – 2024. – № 105. – С. 169–179. [*Dudikhin, V.V., 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. – P. 169–179.] (In Russian)]

26.Стратегическое направление в области цифровой трансформации государственного управления: утв. распоряжением Правительства Российской Федерации от 22.10.2021 № 2998-р. [*Strategicheskoe* napravlenie v oblasti cifrovoj transformacii gosudarstvennogo upravleniya: utv. rasporyazheniem Pravitel'stva Rossijskoj Federacii ot 22.10.2021 № 2998-r. (In Russian)]

27.*Дмитрий* Чернышенко провел стратегическую форсайт-сессию по фундаментальным исследованиям в сфере искусственного интеллекта // Официальный сайт Правительства Российской Федерации. – 30.05.2024. – URL: http://government.ru/news/51726/ (дата обращения 16.02.2026). [*Dmitrii* Chernyshenko provel strategicheskuyu forsait–sessiyu po fundamental′nym issledovaniyam v sfere iskusstvennogo intellekta // Ofitsial′nyy sait Pravitel′stva Rossijskoi Federatsii. – May 30, 2024. – URL: http://government.ru/news/51726/ (accessed February 16, 2026). (In Russian)]

28.*Философия*: Энциклопедический словарь / под ред. А.А. Ивина. – М.: Гардарики, 2004. [*Filosofiya*: Enciklopedicheskij slovar' / pod red. A.A. Ivina. – M.: Gardariki, 2004. (In Russian)]

29.*Философский* энциклопедический словарь / гл. редакция: Л.Ф. Ильичев, П.Н. Федосеев, С.М. Ковалев, В.Г. Панов. – М.: Советская энциклопедия, 1983. [*Filosofskij* enciklopedicheskij slovar' / gl. redakciya: L.F. Il'ichyov, P.N. Fedoseev, S.M. Kovalyov, V.G. Panov. – M.: Sovetskaya enciklopediya, 1983. (In Russian)]

30.*Новая* философская энциклопедия: В 4 т. / под ред. В.С. Степина. – М.: Мысль, 2001. [*Novaya* filosofskaya enciklopediya: Vol. 4 / pod red. V.S. Stepina. – M.: Mysl', 2001. (In Russian)]

31.*Woodward, J.* Scientific Explanation / In: The Stanford Encyclopedia of Philosophy. Ed. by E.N. Zalta. – Stanford: Stanford University, 2017. – URL: https://plato.stanford.edu/archives/fall2023/entries/scientific-explanation/ (дата обращения 15-02-2026). [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.*Закономерность* (закон) // Большая российская энциклопедия: [в 35 т.] / гл. ред. Ю.С. Осипов. – М.: Большая российская энциклопедия, 2004–2017. [*Zakonomernost'* (zakon) // Bol'shaya rossijskaya enciklopediya: [Vol. 35.] / gl. red. YU. S. Osipov. – Moscow: Bol'shaya rossijskaya enciklopediya, 2004–2017. (In Russian)]

34.*Драгалин А. Г.* Математический интуиционизм. Введение в теорию доказательств. – М.: Наука, 1979. – 256 с. – (Математическая логика и основания математики). [*Dragalin, A.G.* Mathematical Intuitionism. Introduction to Proof Theory / Translated by E. Mendelson. – Providence, Rhode Island: AMS, 1988.]

35.*Понимание* / А.А. Ивин, А.Л. Никифоров. Словарь по логике. – М.: Туманит, изд. центр ВЛАДОС, 1997. [*Ponimanie* / A.A. Ivin, A.L. Nikiforov. Slovar' po logike. – M.: Tumanit, izd. centr Vlados, 1997. (In Russian)]

36.*Kitcher, P.* Explanatory Unification and the Causal Structure of the World // In: Scientific Explanation. Ed. by P. Kitcher and W. Salmon. – Minneapolis: University of Minnesota Press, 1989. – P. 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. Ed. by T. Sebeok. – The Hague: Mouton, 1971.

39.*Плотинский Ю.М.* Модели социальных процессов: Учебное пособие для высших учебных заведений. Изд. 2-е, перераб. и доп. – М.: Логос, 2001. – 296 с. [*Plotinskij, Yu.M.* Modeli social'nyh processov: Uchebnoe posobie dlya vysshih uchebnyh zavedenij. Izd. 2 –e, pererab. i dop. – Moscow: Logos, 2001. – 296 s. (In Russian)]

40.*Сохор А. М.* Объяснение в процессе обучения: Элементы дидактической концепции. – М.: Педагогика, 1988. – 128 с. [*Sohor, A.M.* Ob"yasnenie v processe obucheniya: Elementy didakticheskoj koncepcii. – Moscow: Pedagogika, 1988. – 128 s. (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., 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. – P. 347–349. – URL: https://aclanthology.org/W18–5444.pdf/ (дата обращения 15.02.2026). [Accessed February 15, 2026.]

43.*Barskaya, I.* Benchmarks For LLMs // Unite.AI. – August 28, 2024. – URL: https://www.unite.ai/benchmarks-for-llms/ (дата обращения 15.02.2026). [Accessed February 15, 2026.]

44.*Сбер* представил GigaChat 2.0 — новое поколение нейросети. – URL: https://t-j.ru/news/sber-gigachat-2/ (дата обращения 15.02.2026). [*Sber* predstavil GigaChat 2.0 — novoe pokolenie nejroseti. – 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. (дата обращения 15.02.2026). [Accessed February 15, 2026. (In Russian)]

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/ (дата обращения: 16.12.2023). [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., 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. – P. 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. – P. 4762–4779.

58. *Bosselut, A., Le Bras, R., 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. – P. 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

Статья представлена к публикации членом редколлегии академиком РАН Д. А. Новиковым.

Поступила в редакцию 08.07.2025,
после доработки 17.11.2025.
Принята к публикации 16.12.2025.

Кулинич Александр Алексеевич – канд. техн. наук, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва,
✉ alexkul@rambler.ru
ORCID ID: https://orcid.org/0000-0002-4751-205X

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.