

ПРИМЕНЕНИЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ В СИСТЕМАХ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

А.А. Кулинич (*alexkul@rambler.ru*)

Институт проблем управления РАН, Москва

Система поддержки принятия решений представлена как сложная когнитивная система, включающая интеллект лица, принимающего решения в виде модели его ментального пространства и предобученную большую языковую модель. Языковая модель используется для объяснения решений, формально полученных в модели ментального пространства. Предложен метод оценки качества и полезности взаимодействия человека и большой языковой модели. Предложена модель группы респондентов для оценки качества объяснений языковой модели на задаваемые вопросы. Приводится пример оценки языковой модели группой гипотетических респондентов. На примере показан рост понимания сути проблемы лицом, принимающим решение, благодаря объяснениям языковой модели. Предложенная модель системы поддержки принятия решений и метод оценивания качества взаимодействия человека и большой языковой модели могут быть полезны для предварительного оценивания таких систем.

Ключевые слова: поддержка принятия решений, большая языковая модель, объяснение альтернатив решений, оценка качества объяснения решения.

Введение

Для принятия решений в социальных, политических и экономических системах применяют «мягкий» системный анализ [Checkland, 1981]. «Мягкий» системный анализ основывается на принципе ограниченной рациональности [Simon, 1988], который утверждает, что в силу ограниченных когнитивных ресурсов лиц, создающих модель объекта, можно создать упрощенную системную модель с гипотетической структурой, а параметры модели могут иметь лингвистические значения. При построении модели в рамках «мягкого» системного анализа исходными данными являются неструктурированные данные – свободный текст. Модель опи-

сывает объект на ограниченном естественном языке, который предлагает эксперт. Результаты моделирования – альтернативы решений, в такой модели также представлены на ограниченном естественном языке и интерпретируются самим экспертом в терминах его знаний, которых может быть недостаточно для получения нового решения. При этом интерпретация должна включать знания о предметной области релевантные решению, но находящиеся за пределами ограниченного экспертного языка и упрощенной модели ситуации.

В психологии считается, что знания о предметной области представлены в интеллекте человека в психической среде, которая называется ментальным пространством [Fausonnier, 1994] и в котором происходит процесс принятия решения. Это рассуждения, обобщения, интерпретация возможных альтернатив решений. Поддержать процесс принятия решений в ментальном пространстве можно с помощью больших языковых моделей предобученных большим корпусом текста.

В этой работе исследована возможность применения больших языковых моделей в системах поддержки принятия решений в условиях неопределенности для интерпретации и объяснения альтернатив решений.

1. Большие языковые модели в поддержке принятия решений

Применение технологий искусственного интеллекта и машинного обучения в обработке естественного языка привело к появлению больших языковых моделей, популярность которых и области применения расширяются. Большие языковые модели способны генерировать текст, отвечать на сложные вопросы, решать логические задачи, и выполнять множество других задач с качеством сравнимым с качеством выполнения этих задач человеком. Они существенно влияют на многие аспекты нашей жизни: образование, творчество, принятие решений, управление и т.д.

Большие языковые модели (Large Language Model, LLM) 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 [Vaswani, 2017]. Энкодер преобразует входной текст в числовые векторные представления, сохраняя информацию о структуре и взаимосвязях слов. Декодер использует закодированную информацию для генерации ответа или выполнения предсказаний в виде нового оригинального текста, учитывая предыдущие слова и контекст.

Нейронная сеть большой языковой модели представляется в виде многослойной структуры, в которой каждый слой состоит из множества искусственных нейронов, связанных с нейронами соседних слоев. В результате обучения нейронной сети большим корпусом текста формируется словарь токенов – отдельных букв, 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. В результате обучения образуется векторное пространство токенов, в котором определена (косинусная) мера близости токенов. При декодировании, для ответа на вопрос из векторного пространства выбираются токены близкие к теме вопроса, из которых выстраиваются предложения – ответ на вопрос. Выбор токенов из векторного пространства близких к токенам вопроса позволяет сконструировать ответ на поставленный вопрос из слов, часто встречающийся 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, которым обучалась модель. Если в корпус включались тексты из проверенных источников, то ответ будет верен и понятен.

Если языковая модель не понимает вопрос или она обучалась неполными или ошибочными данными, то модель пытается угадать ответ, опираясь на имеющиеся шаблоны, что может привести к ложным ответам – галлюцинациям. Галлюцинация - это феномен больших языковых моделей, когда она генерирует 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 и объяснения, далекие от реальности.

В настоящее время разработаны различные метрики оценки склонности языковых моделей к галлюцинациям и методы их уменьшения. Необходимо отметить, что галлюцинации языковой модели в случаях решения исследовательских или 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 языковых моделей, они находят применение благодаря их способностям решать задачи, представленные на естественном языке, которые ранее решал только человек.

Способности больших языковых моделей оценивают с помощью специальных программ - бенчмарков, которые измеряют их основные качества: объем знаний, точность ответов, надёжность и др. Выделяют бенчмарки: на проверку знаний; логического мышления; понимания прочитанного текста; здравого смысла и др. [Barskaya, 2024].

Бенчмарки на проверку знаний тестируют модели в различных областях. Он оценивает, насколько эффективно модель может вспоминать информацию из разных сфер, таких как физика, география и т.д.

Известный бенчмарк MMLU (Multimodal Language Understanding) создан для проверки уровня знаний модели по различным темам, таким как гуманитарные и социальные науки, история, компьютерные науки и даже право. Он включает 57 вопросов и 15 тысяч задач, направленных на то, чтобы убедиться в уровне знаний 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 тысяч задач по математике для средней школы. Решение этих задач требует выполнения моделью нескольких последовательных элементарных вычислений. Специально обученные языковые модели показывают хорошие результаты на этом бенчмарке, например, модели GPT-4 достигают 96,5%.

Бенчмарк вопросов и ответов (GPQA) уровня аспирантуры оценивает логическое мышление модели, используя набор данных из 448 вопросов. Это сложный тест, разработанный экспертами из областей биологии, физики и химии. Языковая модель GPT-4-omni проходит этот, набирая лишь 53,6%, в то время как аспиранты достигают точности всего 65%.

В 2024 году компания OpenAI выпустила языковую модель OpenAI o1 [Metz, 2024], которая демонстрирует превосходные результаты в сложных рассуждениях, превосходя людей в тестах по математике, кодированию и естественным наукам. На отборочных экзаменах Международной математической олимпиады (IMO) новая модель правильно решила 83% задач, тогда как ее предшественница GPT-4o дала лишь 13% верных ответов. Утверждается, 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 ответы. Тестирование заключается в ответах на вопросы по текстам, что позволяет оценить понимание, способность делать выводы и удерживать 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 96,000 вопросов для проверки способностей языковой модели к рассуждению. Вопросы DROP содержат информацию, которая требует от моделей выполнения математи-

ческих операций, на основе информации, разбросанной по всему тексту. Модели GPT-4 достигает точности 80%, в то время как люди достигают 96% правильных ответа на наборе тестов DROP.

Бенчмарки на проверку общих знаний (здорового смысла) оценивают способность модели к обобщённым 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 заданного сценария. Тестирование методами HellaSwag показывают, что современные модели, такие как GPT-4, достигли уровня производительности, близкого к человеческой точности.

Тестирование способностей языковых моделей с помощью бенчмарков позволяет объективно оценить и сравнить качество моделей разных производителей, понять текущий уровень и динамику развития языковых моделей. Как показал анализ тестирования способностей языковых моделей на тестах бенчмарков, эти способности приближаются к способностям человека. Это означает, что большая языковая модель может быть ассистентом в системах поддержки принятия решений.

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

В этом случае целесообразно для оценки привлекать экспертов, способных дать точную и надежную оценку способностей больших языковых моделей как ассистентов в поддержке принятия решений.

2. Семиотическая модель ментального пространства и генерация альтернатив решений

В работе [Kulinich, 2023a] были рассмотрены вопросы генерации и интерпретации решений в условиях ограниченной рациональности с помощью семиотической модели ментального пространства. Семиотическая модель ментального пространства – это модель знаний эксперта о ситуации принятия решений, которая включает три взаимодействующие модели: синтаксическая, семантическая и прагматическая модели. В синтаксиче-

ской модели на ограниченном естественном языке эксперт описывает ситуацию принятия решений, используя имена всех объектов и их параметров, множество их возможных лингвистических значений, причинно-следственные отношения на множествах значений. Причинно-следственные отношения задают модели рассуждений эксперта в виде причинной сети. Рассуждения эксперта заключается в изменении им значений параметров, которые по причинным связям передаются другим параметрам ситуации принятия решений. В результате изменяются все связанные параметры ситуации. Новый вектор значений параметров, полученный в результате рассуждения эксперта, передается в семантическую модель. Семантическая модель представляется в виде качественного концептуального каркаса. Это качественная идеализированная онтология предметной области, в которой задана начальная ситуация, ее имя, параметры, их значения и все возможные обобщения начальной ситуации по разным параметрам. Начальная ситуация и обобщенные ситуации связаны родовидовыми отношениями (класс-подкласс) образуя решетку понятий предметной области. Рассуждая в синтаксической системе, эксперт может перейти к обобщенным ситуациям, что приводит к появлению нового класса решений и новым альтернативам.

В семантической модели мы получаем решения в виде составных имен классов решений, которые выражены на ограниченном естественном языке автора этой модели. Для представления решения в альтернативных языковых формах решения интерпретируются и объясняются с помощью больших языковых моделей языка.

3. Интерпретации и оценка объяснения языковыми моделями альтернатив решений

В работе [Kulinich, 2023b] было предложено интерпретировать и объяснять составные имена классов решений с помощью статических и динамических векторных моделей языка, представленных искусственной нейронной сетью. В работе [Hoffman, 2018] приводятся системные характеристики системы объяснения.

Во-первых, система объяснения – это сложная когнитивная система, включающая интеллект человека и обученную языковую модель. Со стороны человека в формировании объяснений задействовано его ментальное пространство. Оно отвечает за понимание ответов сгенерированных языковой моделью. Если объяснение языковой модели понятно человеку, то оно увеличивает его ментальное пространство.

Во-вторых, система объяснения – это диалоговая система. Если языковая модель отвечает на вопрос, то у человека могут возникнуть дополнительные вопросы. В этом случае хорошим вариантом дополнительных вопросов является метод «5 почему?» [Sarrat, 2017]. Он позволяет исследовать причинно-следственные цепочки проблемы.

В третьих, система объяснения – это исследовательская система. Она должна поддерживать любопытство человека, заставляя его разобраться в проблеме и найти нужное решение.

В работе [Hoffman, 2018] предложен ряд критериев для оценки качества систем объяснения. Это критерии: Полезность объяснения; Удовлетворенность объяснением; Понятность объяснения; Способность объяснения стимулировать любопытство; Доверие к объяснению.

Перечисленные критерии можно измерить. В работе [Hoffman, 2018] предложено для определения значений критериев в этой когнитивной системе, включающей интеллект человека (его ментальное пространство) и обученную нейронную сеть, использовать метод тестирования группой респондентов объяснений языковой модели по каждому из перечисленных критериев. Задача тестирования заключается в измерении латентной переменной – это удовлетворенность респондентов качеством объяснения языковой моделью альтернатив решений и контекста их применения.

Серьезной проблемой такого подхода является формирование группы респондентов. Формирование группы респондентов для измерения латентной переменной отношения человека к объяснению языковой модели считается искусством. В этой работе для предварительной оценки латентной переменной предложена модель группы респондентов.

4. Модель группы респондентов

Для тестирования каждого критерия используется семизначная шкала Лайкерта [Likert, 1932]. Пример шкалы Лайкерта приведен в табл. 1.

Модель респондента это пара - (E, σ) , где E – оценка респондента, $E \in \{1, 2, 3, 4, 5, 6, 7\}$, σ - среднее отклонение оценки эксперта.

Будем различать три типа респондентов: (Эксперт (E, σ_1) , Эрудит (E, σ_2) , Новичок (E, σ_3)). Эксперт всегда дает оценку с малым средним отклонением. Экспертная оценка знатока, характеризуется большим средним отклонением, и наконец, новичок имеет самое большое среднее отклонение, т.е. $\sigma_1 < \sigma_2 < \sigma_3$.

Группа респондентов формируется из респондентов разных типов. Число респондентов разных типов является параметром модели и может варьироваться в широком диапазоне. Далее считаем, что известна оценка одного эксперта объяснения языковой модели на вопрос «почему?» по всем пяти критериям, $E = (e_1, e_2, e_3, e_4, e_5)$, где $e_i \in \{1, 2, 3, 4, 5, 6, 7\}$. Считаем, что в группе N_1 экспертов, N_2 знатоков и N_3 новичков.

Для каждого эксперта сформируем случайную оценку $E_{ex} = (e_1 \pm \sigma_1, e_2 \pm \sigma_1, e_3 \pm \sigma_1, e_4 \pm \sigma_1, e_5 \pm \sigma_1)$, $ex = 1, \dots, N_1$. Аналогично, сформируем оценки для знатоков, $E_{co} = (e_1 \pm \sigma_2, e_2 \pm \sigma_2, e_3 \pm \sigma_2, e_4 \pm \sigma_2, e_5 \pm \sigma_2)$, $co = 1, \dots, N_2$ и новичков, $E_{be} = (e_1 \pm \sigma_3, e_2 \pm \sigma_3, e_3 \pm \sigma_3, e_4 \pm \sigma_3, e_5 \pm \sigma_3)$, $be = 1, \dots, N_3$. Число из диапа-

зон отклонений $\pm\sigma_1, \pm\sigma_2, \pm\sigma_3$ – формирует генератор случайных чисел для каждого респондента. Таким образом, формируется случайная анкета, содержащая оценки $N_1 + N_2 + N_3$ респондентов.

Таблица 1

Шкала Лайкерта

Полез- ность	Удовлетво- ренность	Понятность	Любопыт- ство	Доверие	Оцен- ка
Очень полезно	Очень доволен	Абсолютно понятно	Обязатель- но возни- кает	Абсолютно доверяю	7
Полезно	Доволен	Понятно	Возникает	Доверяю	6
Мало полезно	Мало дово- лен	Мало понятно	Мало возникает	Мало доверяю	5
Затрудня- юсь отве- тить	Затрудняюсь ответить	Затрудняюсь ответить	Затрудня- юсь отве- тить	Затрудня- юсь отве- тить	4
Скорее бесполезно	Скорее недо- волен	Скорее не понятно	Скорее не возникает	Скорее не доверяю	3
Бесполезно	Недоволен	Не понятно	Не возни- кает	Не доверяю	2
Совершен- но беспо- лезно	Совершенно недоволен	Совершенно непонятно	Совершен- но не воз- никает	Совершен- но не дове- ряю	1

Однако, такая случайная анкета не может быть использована для оценки латентной переменной, характеризующей отношение лица, принимающего решение к объяснениям языковой модели.

Для проверки согласованности случайной анкета воспользуемся показателем α -Кронбаха [Cronbach, 1951]:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_x^2} \right) \quad (1)$$

где k – число вопросов респондентам, σ_i^2 – среднее квадратичное отклонение оценок респондента по каждому вопросу, σ_x^2 – среднее квадратичное отклонение суммы оценок респондентов.

Если показатель $\alpha > 0.7$, то считается, что случайная анкета согласована. Если анкета не согласована, то осуществляется новая генерация случайной анкеты до получения согласованной анкеты.

Для согласованных анкет рассчитываем коэффициент ранговой корреляции Спирмена по формуле 2:

$$r = 1 - 6 \frac{\sum_{i=1}^n d_i^2}{n^3 - n} \quad (2)$$

где n – число респондентов, d_i^2 – квадрат разности рангов оценок респондентов по каждому вопросу и рангов суммы этих оценок.

Ранговая корреляция рассчитывается для каждого критерия, т.е. получаем вектор корреляций $R=(r_1, r_2, r_3, r_4, r_5)$. Для каждого коэффициента корреляции определим его статистическую значимость. По следующей формуле рассчитываем наблюдаемое значение t -критерия:

$$t_{obs} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \quad (3)$$

где n – число респондентов, r – значение корреляции, $r \in R$.

По таблице критических значений Стьюдента для значения n определяем критическое значение t_{cr} . Если наблюдаемый t -критерий больше критического, $t_{obs} > t_{cr}$, то корреляция значима, т.е. гипотезу о существовании корреляции можно принять. Статистическая значимость коэффициента корреляции по некоторому вопросу позволяет говорить о нормальности распределения ответов респондентов по этому вопросу. Это повышает доверие к случайно сформированной анкете, основанной на одной экспертной оценке.

Такой подход может быть полезен для качественной оценки различных больших языковых моделей в системах принятия решений.

5. Пример

Большой языковой модели «`rugpt_medium_turbo_instructed`» [Huggingface, 2025] задавался ряд вопросов по методу «5 почему?»

Первый вопрос к языковой модели такой: *Почему у бизнесменов возникают финансовые проблемы?*

Ответ языковой модели: *Финансовые проблемы могут возникнуть в различных ситуациях, включая отсутствие прибыли и роста бизнеса. Некоторые из наиболее распространенных причин - это потеря денег на инвестициях или кредитах, ухудшение финансового положения компании, снижение эффективности работы сотрудников, изменение законодательства о налогах и сборах, а также другие факторы.*

Второй вопрос: *Почему у бизнесменов нет прибыли и роста бизнеса?*

Ответ языковой модели: *Бизнесмены не могут получать прибыль в связи с отсутствием доходов. Причина - отсутствие прибылей, которые позволили бы им развиваться дальше.*

Третий вопрос: *В каких условиях должны работать бизнесмены, чтобы были доходы?*

Ответ языковой модели: *Бизнесмен должен иметь разные источники дохода – от продажи товаров и услуг до управления бизнесом в целом. Однако для того, чтобы стать успешным бизнесменом, необходимо уметь управлять своими финансами, инвестировать деньги в акции или облигации, а также работать эффективно.*

Для эксперта была разработана анкета, в которой предлагалось последовательно оценить все три вопроса и ответы языковой модели на них. Вопросы были следующие:

- Оцените полезность этого объяснения?
- Вы удовлетворены следующим объяснением?
- Вам понятно это объяснение?
- Возникают при таком объяснении дополнительные вопросы?
- Вы доверяете такому объяснению?

Вопросы предназначены для экспертной оценки каждого ответа языковой модели по критериям полезности, удовлетворённости, понятности, способности стимулировать любопытство и доверию к объяснению. Оценки по всем вопросам и объяснениям языковой модели с использованием шкалы Лайкерта приведены в табл. 2.

Таблица 2

Оценки Эксперта

	Полезность	Удовлетворенность	Понятность	Любопытство	Доверие
Вопрос 1	Полезно (6)	Доволен (6)	Понятно (6)	Скорее не возникает (3)	Доверяю (6)
Вопрос 2	Полезно (6)	Мало доволен (5)	Понятно (6)	Обязательно возникает (7)	Доверяю (6)
Вопрос 3	Малополезно (5)	Мало доволен (5)	Малопонятно (5)	Обязательно возникает (7)	Мало доверяю (5)

Векторы оценок эксперта, следующие по: вопросу 1 –(6, 6, 6, 3, 6); вопросу 2 –(6, 5, 6, 7, 6); вопросу 3 –(5, 5, 5, 7, 5).

Формировалась случайная анкета из ответов респондентов, для которой рассчитывался коэффициент α -Кронбаха (1). Многочисленные генерации случайных анкет позволили выделить по 12 анкет для каждого вопроса с удовлетворительным значением коэффициента согласованности.

Коэффициенты ранговой корреляции рассчитывались по формуле (2). Значения усредненных коэффициентов ранговой корреляции приведены в табл. 3.

Таблица 3

Ранговая корреляция для критериев оценивания

	Вопрос 1	Вопрос 2	Вопрос 3
Полезность	0,68	0,69	0,75
Удовлетворенность	0,66	0,65	0,78
Понятность	0,48	0,57	0,63
Любопытство	0,68	0,57	0,59
Доверие	0,67	0,72	0,73

Наблюдаемые значения t-критерия рассчитывались по формуле (3). Критические значения t-критерия выбирались из таблицы критических значений Стьюдента для значений $\beta=0.05$.

В результате проверки для $\beta=0.05$ коэффициенты корреляции по всем критериям оказались статистически значимы. Это означает, что вероятность того, что эти коэффициенты корреляции случайные составляет всего 5 процентов.

Анализ динамики изменения коэффициентов корреляции по разным критериям по таблице 3 показывает, что при последовательной оценке вопросов по методике «5 почему?» происходит рост значений коэффициентов по всем критериям, кроме критерия «Любопытство». Это может свидетельствовать о росте ментального пространства респондента. То есть по мере того как он получает ответы языковой модели, он лучше понимает проблему и меньше задает дополнительных вопросов.

Заключение

Рассмотрена модель сложной когнитивной системы, включающей интеллект лица, принимающего решения в виде модели его ментального пространства и большой языковой моделью. Большая языковая модель используется для интерпретации и объяснения решений, формально полученных в модели ментального пространства.

Предложен метод оценки качества и полезности взаимодействия человека и большой языковой модели. Рассмотрена модель гипотетической группы респондентов для оценки качества объяснений языковой модели на задаваемые вопросы. Были выполнены эксперименты с группой гипотетических респондентов, которые показали рост понимания сути проблемы лицом, принимающим решение благодаря объяснениям языковой модели.

Предложенная модель системы поддержки принятия решений, включающая большие языковые модели и метод оценивания качества взаимодействия человека и большой языковой модели может быть полезен для предварительного оценивания таких систем.

Список литературы

- [Barskaya, 2024] Barskaya I. Benchmarks For LLMs. Unite.AI, Aug. 28, 2024. – <https://www.unite.ai/benchmarks-for-llms/>.
- [Checkland, 1981] Checkland P.B. Systems Thinking, Systems Practice. – New York: Wiley, 1981. – 330 p.
- [Cronbach, 1951] Cronbach L.J. Coefficient alpha and the internal structure of tests // Psychometrika. – Springer Science and Business Media LLC, 1951. – 16(3). – P. 297-334. – doi:10.1007/bf02310555. – hdl:10983/2196. S2CID 13820448.
- [Fauconnier, 1994] Fauconnier G. Mental Spaces: Aspects of meaning construction in natural language. – Cambridge: Cambridge University Press, 1994.
- [Hoffman, 2018] Hoffman R., Mueller S., Klein G., Litman J. Metrics for explainable AI: Challenges and prospects // arXiv:1812.04608. – 2018.
- [Huggingface, 2025] <https://huggingface.co/models/>.
- [Kulinich, 2023a] Kulinich A. A. Semiotic Model of Mental Space // Pattern Recognit. Image Anal. – 2023. – Vol. 33. – P. 373-382. – <https://doi.org/10.1134/S1054661823030239>.
- [Kulinich, 2023b] Kulinich A.A. A Model to Support Intuitive Decision-Making // Proceedings of the 16th International Conference Management of Large-Scale System Development (MLSD). Moscow, Russian Federation: IEEE Xplore, 2023. – P. 1-5. – <https://ieeexplore.ieee.org/document/10304012>.
- [Likert, 1932] Likert R. A technique for the measurement of attitudes // Arch. Psychol. – 1932. – Vol. 7, No. 140.
- [Metz, 2024] Metz C. OpenAI Unveils New ChatGPT That Can Reason Through Math and Science. The New York Times (12 sept. 2024) (дата обращения: 28-05-2025).
- [Sarrat, 2017] Sarrat, O. The Five Whys Technique. In. Knowledge Solutions. – Springer, Singapore, 2017. – https://doi.org/10.1007/978-981-100983-9_32.
- [Simon, 1988] Simon HA. Rationality as Process and as Product of Thought / In: Bell DE, Raiffa H, Tversky A, eds. Decision Making: Descriptive, Normative, and Prescriptive Interactions. – Cambridge University Press, 1988. – P. 58-77.
- [Vaswani, 2017] Vaswani A., Shazeer N., et. all. Attention is all you need // arXiv pre-print arXiv:1706.03762, 2017. – <https://doi.org/10.48550/arXiv.1706.03762>.