



ПРИМЕНЕНИЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ В СИСТЕМАХ ПОДДЕРЖКИ ПРИНЯТИИ РЕШЕНИЙ.

Ч. 2. Измерение и оценка удовлетворенности лица, принимающего решение

А. А. Кулинич

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

✉ alexkul@rambler.ru

Аннотация. Во второй части исследования, посвященного применению больших языковых моделей в системах поддержки принятия решений, рассматриваются вопросы измерения качества гибридной системы поддержки принятия решений, включающей лицо, принимающее решение, и в качестве ассистента большую языковую модель. Предложены множество критериев оценки качества гибридной системы поддержки принятия решений и экспертная процедура оценивания. Особенностью экспертной оценки качества системы является предложенная структурная модель разумного эксперта, на основе которой осуществляется квазиэксперимент оценивания гибридной системы. Оценивание осуществляется в рамках методики выявления причин проблемы «пять почему?». Приводятся пример работы и оценивания гибридной системы поддержки принятия решений и анализ результатов эксперимента. Предложенный метод измерения качества и оценки удовлетворенности лица, принимающего решение, основанный на структурной модели разумного эксперта и квазиэксперимента оценивания, может быть полезен на этапе создания системы поддержки принятия решений, включающей большую языковую модель. Этот метод позволит получить качественные оценки гибридных систем поддержки принятия решений и сократить время и стоимость оценивания без привлечения групп экспертов.

Ключевые слова: поддержка принятия решений, большая языковая модель, гибридная система, критерии оценивания, структурная модель разумного эксперта.

ВВЕДЕНИЕ

В первой части исследования были рассмотрены вопросы применения в системах поддержки принятия решений больших языковых моделей в качестве объясняющего ассистента в процессах исследования ситуации принятия решения, генерации альтернатив решений и выбора, обоснования решения и его реализация [1]. Во второй части статьи речь пойдет об оценке взаимодействия человека и большой языковой модели (искусственного интеллекта). В данном случае речь идет о человеко-машинном взаимодействии и оценке его качества.

В литературе, когда речь заходит о человеко-машинном взаимодействии, подразумевается интерфейс машины (ЭВМ) и пользователя. Назначение такого интерфейса, его задачи – это повышение производительности труда оператора, повышение надежности решения задачи и др. При этом качество интерфейсов оценивают с точки зрения качества выполнения вышеприведенных задач для конкретных областей применения. Для повышения качества интерфейсов, как правило, опираются на психологические особенности восприятия человека, его моторные характеристики и др. [2].

При оценивании качества взаимодействия человека с большой языковой моделью есть некото-

рые особенности. Во-первых, это нулевой интерфейс. Это означает, что человек общается с компьютером на естественном языке – например, голосом или через стандартную эргономичную клавиатуру. Во-вторых, языковая модель дает ответ на вопрос на естественном языке (это не стрелочный прибор или какой-то другой измеритель) и здесь также должны быть задействованы психологические особенности человека, но это не особенности восприятия или моторные характеристики, а механизмы мышления, понимания, аргументации и др.

В литературе есть работы, посвященные оценке качества объяснений, которые выдают системы искусственного интеллекта. Это так называемые системы объяснимого искусственного интеллекта (англ. *eXplainable AI*, XAI) [3]. Эти системы дают объяснение вывода нейронной сетью ответа на вопрос. Объяснение вывода направлено на повышение доверия к системе искусственного интеллекта и исключение ошибок при принятии ответственных решений. Такие объяснения включают причинно-следственные цепочки от запроса до результата в терминах понятий предметной области, в которой нейронная сеть была обучена. В подобном объяснении нет семантики предметной области, есть только логический вывод.

Получение такого вывода в большой языковой модели сопряжено с некоторыми проблемами. Большая языковая модель – это нейронная сеть глубокого обучения. Известно, что сети глубокого обучения – это сети представления данных. Это означает, что обученная нейронная сеть представляет структуру данных, содержащуюся в обучающей выборке. Большие языковые модели обучаются большим корпусом текста, а встроенные механизмы внимания выделяют синтаксическую структуру текста [4]. То есть применять методы извлечения причинно-следственных цепочек вывода результата для понимания его корректности здесь бессмысленно. Вывод будет говорить о синтаксической структуре предложения, а не об особенностях предметной области.

Возможности языковой модели в обработке естественного языка, понимания, рассуждения оценивают с помощью тестовых программ – бенчмарков. С помощью тех же бенчмарков можно оценить способности человека решать сложные задачи, представленные на естественном языке, и сравнить результаты решения задачи человеком с результатами языковой модели [5]. В этом случае способности человека и языковой модели оцениваются с помощью одних и тех же тестов отдельно, а сравнение результатов позволяет говорить об

успехах разработчиков языковой модели, о приближении способностей языковой модели к способностям человека.

Бенчмарки тестируют отдельные способности языковой модели. На поставленный тестовый вопрос, как правило, бенчмарки имеют однозначный ответ. Это дает возможность измерить отдельную способность. При принятии решений в сложных социальных и экономических ситуациях все эти способности работают одновременно в некоторых сочетаниях. Кроме того, решение в таких сложных ситуациях не единственное – это множество возможных решений, выбрать лучшее из которых оказывается затруднительным: все либо проверяется на практике, либо требует дополнительной оценки и обоснования. В таких ситуациях обычные бенчмарки не работают, нужна экспертная оценка.

Особенность такой экспертной оценки заключается в том, что необходимо оценить взаимодействие человека и большой языковой модели, которое нужно рассматривать как сложную когнитивную систему, функционирование которой направлено на получение качественных решений по управлению сложными ситуациями. В этом случае языковая модель выдает статистически правдоподобный ответ (наиболее вероятный ответ, включенный в обучающую выборку), не понимая его смысла, а человек пытается понять ответ и встроить его в собственную систему знаний. Понимание человеком ответа языковой модели в значительной степени зависит от уровня его знаний, и поэтому разные люди, получая один и тот же ответ языковой модели, могут дать ему разные оценки. Система оценки объяснения в этом случае должна быть индивидуальна.

Таким образом, задача заключается в разработке метода индивидуальной оценки удовлетворенности человека объяснениями ситуации принятия решения, которые он получает от большой языковой модели.

В первой части статьи [1] были сформулированы цель и задачи объяснения языковой модели в условиях неопределенности. Во второй части будет рассмотрен метод оценки удовлетворенности лица, принимающего решение (ЛПР), объяснением для решения исследовательской и практической задач принятия решения. По результатам экспериментов с объяснениями реальных больших языковых моделей будет рассмотрен вариант классификации этих объяснений в классы моделей объяснений, рассмотренных в первой части статьи [1].



1. КРИТЕРИИ КАЧЕСТВА ОБЪЯСНЕНИЯ СИСТЕМЫ ЧЕЛОВЕК – ЯЗЫКОВАЯ МОДЕЛЬ

В работе [3] приводятся системные характеристики системы объяснения.

Во-первых, система объяснения – это сложная когнитивная система, включающая человека, его интеллект и обученную нейронную сеть. Со стороны человека в функционировании объяснений задействовано его ментальное пространство. Оно отвечает за понимание ответов, сгенерированных языковой моделью. Считается, что если объяснение языковой модели понятно человеку, то происходит увеличение его ментального пространства.

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 «пять почему?» [6]. Этот метод позволяет построить причинно-следственные цепочки проблемы и выяснить ее причину.

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

В работах [7, 8] на бенчмарках исследуются промпты здравого смысла для больших языковых моделей. Кроме повышения качества работы языковой модели в результате применения этих промптов, в указанных работах исследуется человеческий фактор, включенный в когнитивную систему объяснения. В частности, отмечается, что некоторые корректные (согласно бенчмарку) объяснения языковой модели эксперты оценивают как некорректные, а соответственно, некоторые некорректные как корректные. При этом некоторые корректные ответы оцениваются как совершенно бесполезные. Получается, что качество когнитивной системы поддержки принятия решений, включающей человека и языковую модель, в значительной степени зависит от уровня знаний человека и определяется им самим.

Для оценки качества такой когнитивной системы в работах [7, 8] применяется экспертный опрос по следующим критериям:

- Грамотность: является ли объяснение грамматически правильным (значения шкалы: грамматически правильный; неграмотный, но понятный; совершенно бессмысленный).

- Релевантность: соответствует ли объяснение теме, упомянутой в вопросе (значения шкалы: релевантный; нерелевантный).

- Достоверность (фактичность): является ли объяснение фактически верным (значения шкалы: достоверно; недостоверно).

- Полезность: помогает ли объяснение ответить на вопрос напрямую или косвенно (значения шкалы: полезное; нейтральное; вредное).

Как видим, авторы предлагают использовать номинальную шкалу для оценивания критериев. Для работы с такими шкалами требуется большое количество респондентов и специфические методы обработки результата опроса. Судя по работам [7, 8], оценивание проводилось для небольшого числа респондентов и удалось получить приемлемую согласованность ответов.

В публикации [7] при оценивании работы промпта, основанного на генерации подсказки здравого смысла с помощью испытуемой языковой модели, 82 % объяснений оказались грамматически правильными и достоверными, при этом только 72 % полезными. В работе [8] при оценивании промпта, основанного на внутреннем диалоге с языковой моделью, из множества достоверных ответов только 40 % оказались полезными.

Вывод, который можно сделать по результатам обработки опроса респондентов говорит о важности задачи оценивания когнитивной системы «человек и языковая модель».

В работе [3] предложен ряд критериев для оценки качества систем объяснения, в частности:

- полезность объяснения,
- понятность объяснения,
- реализуемость объяснения,
- доверие к объяснению,
- способность объяснения стимулировать любопытство.

Перечисленные критерии можно измерить. В работе [3] предложено для определения значений критериев в этой сложной когнитивной системе применять метод тестирования группой респондентов объяснений нейросети по каждому из перечисленных критериев с использованием порядковой шкалы Лайкерта [9].

Задача тестирования заключается в измерении латентной переменной – мнения респондентов о качестве объяснения языковой моделью закономерностей предметной области и по возможности оценки расширения ментального пространства человека благодаря объяснениям языковой модели. В дальнейшем будем опираться на эту систему критериев, ссылаясь на авторитет ее авторов как известных специалистов в области оценки человеко-машинного взаимодействия.

2. ИЗМЕРЕНИЕ ЛАТЕНТНЫХ ПЕРЕМЕННЫХ

Традиционным подходом в социологии и психологии для измерения латентных переменных является экспертный опрос. Под латентной переменной понимается отношение человека или социальной группы к некоторому явлению, непосредственное измерение которого невозможно. Латентную переменную можно определить через некоторые переменные, которые предположительно связаны с латентной переменной и которые можно измерить в порядковых или интервальных шкалах. Методы математической статистики и теории измерений позволяют определить значение латентной переменной в порядковой или интервальной шкале.

В социологии для измерения латентной переменной разработано большое количество методов и приемов, на которых пока останавливаться не будем, поясним лишь суть измерения. Для измерения латентной переменной путем измерения наблюдаемых и измеряемых переменных исследователь разрабатывает специальные опросные анкеты по каждой измеряемой переменной с возможными вариантами ответов. Анкеты раздаются множеству респондентов, которые отвечают на эти вопросы. Далее различными методами математической статистики результат опроса обрабатывается и получается некоторая оценка латентной переменной. Желательно, чтобы оценка латентной переменной была в интервальной шкале. В этом случае появляется возможность сравнивать значения латентных переменных множества измеряемых объектов и давать числовые оценки такому сравнению.

Проблема заключается в получении оценки латентной переменной в интервальной шкале. Для этого необходимо, чтобы группа респондентов была однородна. Для формирования группы респондентов проводятся специальные исследования. Если группа подобрана, то и в этом случае возможны неоднородности опроса, которые устраняются исследователем вручную. Эта процедура называется ремонтом выборки, она зачастую приводит к формированию новой группы респондентов, изменению анкеты вопросов или вариантов ответов. Опрос проводится повторно до получения приемлемых результатов. Отметим, что организация опроса включает в себя формирование группы, разработку анкеты, обработку результатов, возможно, ремонт выборки или повторный опрос; все это требует больших временных и материальных затрат.

3. МОДЕЛЬ РАЗУМНОГО РЕСПОНДЕНТА И ГРУППЫ РЕСПОНДЕНТОВ

Специфика предлагаемой оценки качества объяснения языковой модели в процессах поддержки принятия решений заключается в том, что понимание человеком ответа языковой модели в значительной степени зависит от уровня его знаний, и поэтому разные люди, получая одни и те же объяснения языковой модели, могут дать ему разные оценки. В этом случае модель оценки должна быть индивидуальна. Групповые оценки удовлетворенности человека объяснением языковой модели могут быть неточны.

В настоящее время популярен так называемый модельный подход в социологии [10]. С развитием технологии искусственного интеллекта формируется новое научное направление в исследовании социальных систем – вычислительная социология. Это междисциплинарная область, использующая методы вычислительной науки для изучения социальных явлений. Она применяет вычислительные методы и анализ больших данных для разработки и проверки теорий о социальных системах, моделях поведения и взаимодействий [11].

В настоящей работе предлагается построить модель разумного респондента, представить множество разумных респондентов как многоагентную систему, моделирующую однородную группу респондентов, и далее измерить латентную переменную с помощью традиционных в социологии методов.

Параметры модели разумного респондента зависят от множества измеряемых переменных, особенностей предметной области, в которой строится модель, и от предпочтений ЛПР. Поэтому далее будем строить модель для ранее определенных критериев оценки удовлетворенности ЛПР объяснениями. Напомним, что в качестве критериев выступают полезность объяснения, понятность объяснения, реализуемость объяснения, способность объяснения стимулировать любопытство, доверие к объяснению.

В модельном подходе в социологии модель должна отражать закономерности поведения респондента в процессах оценивания измеряемых переменных. Эти закономерности поведения описываются с помощью структурных уравнений – зависимостей значений измеряемых переменных [12]. Как правило, в общем виде – это линейные уравнения вида $Y = X + \tilde{x}$, где Y – латентная переменная; X – измеряемая переменная; \tilde{x} – случайная ошибка измерения. Если число измеряемых



переменных велико, то для наглядности зависимости могут быть представлены в виде ориентированного графа. Пример такого графа для рассматриваемых критериев представлен на рис. 1.

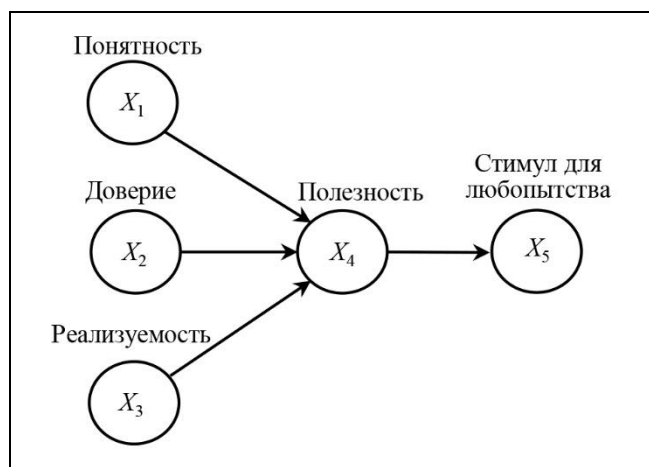


Рис. 1. Структурная модель разумного респондента

Структурные уравнения, реализующие поведение разумного респондента для рассматриваемого случая имеют вид неравенств

$$X_4 \leq X_1 + x_1^{\sim}, \tag{1}$$

$$X_4 \leq X_2 + x_2^{\sim}, \tag{2}$$

$$X_4 \leq X_3 + x_3^{\sim}, \tag{3}$$

$$X_5 \geq X_4 + x_4^{\sim}, \tag{4}$$

где X_1, X_2, X_3, X_4 – измеряемые переменные; $x_1^{\sim}, x_2^{\sim}, x_3^{\sim}, x_4^{\sim}$ – случайные ошибки измерения переменных.

Значения этих переменных зависят от способа их измерения. Далее определим порядковую шка-

лу Лайкерта, с помощью которой будем измерять значения переменных и, следовательно, определим их возможные значения (табл. 1).

Поясним неравенства (1)–(4). В них заложены закономерности разумного поведения респондента. Из неравенства (1) следует, что объяснение не может быть полезным, если оно непонятно. Из неравенства (2) следует, что объяснение не может быть полезным, если ему нельзя доверять. Из неравенства (3) следует, что объяснение не может быть полезным, если его нельзя реализовать. Из неравенства (4) следует, что если объяснение полезно, то стимул для любопытства падает. Это означает, что объяснение «встроилось» в систему знаний человека, увеличив их, и в дальнейших исследованиях необходимости нет.

Необходимо отметить, что модель разумного респондента строится самим потребителем объяснения языковой модели, т. е. она субъективна. Потребитель выбирает критерии – измеряемые переменные, определяет их зависимости, а это значит, что приведенная выше модель может быть перестроена субъектом под свои представления. Отметим также, что построенная субъективная модель может отражать поведение респондентов – единомышленников ЛПР, который построил модель. Под единомышленниками здесь понимается группа респондентов, которые разделяют общее мнение, убеждения и имеют схожее поведение в процессах оценки критериев в схожих ситуациях принятия решения. Если опросить такую группу респондентов, то результат такого опроса будет однородным и его можно использовать для оценки удовлетворенности объяснениями языковой модели.

Гипотеза об однородности группы респондентов, действующих согласно модели разумного поведения при оценивании, проверялась.

Таблица 1

Шкала Лайкерта

Полезность	Реализуемость	Понятность	Любопытство	Доверие	Оценка
Очень полезно	Легко реализуется	Абсолютно понятно	Обязательно возникает	Абсолютно доверяю	7
Полезно	Реализуется	Понятно	Возникает	Доверяю	6
Малополезно	Трудно реализуется	Малопонятно	Слабо возникает	Слабо доверяю	5
Затрудняюсь ответить	Затрудняюсь ответить	Затрудняюсь ответить	Затрудняюсь ответить	Затрудняюсь ответить	4
Скорее бесполезно	Скорее не реализуется	Скорее непонятно	Скорее не возникает	Скорее не доверяю	3
Бесполезно	Не реализуется	Непонятно	Не возникает	Не доверяю	2
Совершенно бесполезно	Невозможно реализовать	Совершенно непонятно	Совершенно не возникает	Совершенно не доверяю	1

Для оценки значений измеряемых переменных использовался популярный в социологии метод Лайкерта [9]. Шкала Лайкерта приведена в табл. 1.

В столбцах табл. 1 представлены названия критериев и их возможные лингвистические значения. Лингвистическим значениям соответствуют числа от 1 до 7, шкала порядковая. Оценивая объяснение языковой модели, респондент отвечает на вопросы по пяти критериям, проставляя лингвистические оценки.

Его оценки представляются в виде числового вектора $R = (r_1, r_2, r_3, r_4, r_5)$, $r_i \in \{1, 2, 3, 4, 5, 6, 7\}$, при этом все оценки должны удовлетворять модели разумного респондента.

В работах известного психофизика и одного из основоположников теории измерений в социологии Л. Терстоуна выдвигается интересная гипотеза [13]. Утверждается, что если одного и того же респондента в разные моменты времени попросить дать оценку какому-либо явлению или объекту, то эти разнесенные по времени оценки будут подчиняться закону нормального распределения с определенными математическим ожиданием и дисперсией. В тоже время, если одновременно попросить дать оценку тому же явлению или объекту группу респондентов, то их оценки также будут распределены по нормальному закону с определенными математическим ожиданием и дисперсией. При определенных условиях нормальные распределения одного респондента и группы могут совпасть. Например, если группа респондентов однородна и все они единомышленники респондента, давшего множество разнесенных во времени оценок. На этом предположении и ряде других допущений у Л. Терстоуна строится строгое математическое доказательство возможности измерения латентной переменной [13, 10].

Пусть в результате опроса множества респондентов G было получено множество оценок наблюдаемого явления или объекта $R_j = (r_{1j}, r_{2j}, r_{3j}, r_{4j}, r_{5j})$, где $r_{1j}, r_{2j}, r_{3j}, r_{4j}, r_{5j}$ – оценки j -го респондента по всем пяти критериям из множества оценок шкалы Лайкерта $r_{ij} \in \{1, 2, 3, 4, 5, 6, 7\}$; $i \in \{1, 2, 3, 4, 5\}$ – номер критерия, $R_j \in G$.

Определим математическое ожидание и среднеквадратическое отклонение по каждому критерию для группы респондентов G как вектор пар $(M(r_1), \sigma_1; M(r_2), \sigma_2; M(r_3), \sigma_3; M(r_4), \sigma_4; M(r_5), \sigma_5)$, где $M(r_i)$ – математическое ожидание оценки группы респондентов по i -му критерию; σ_{ij} – среднеквадратическое отклонение соответствующей оценки.

Считается, что вектор математических ожиданий $(M(r_1), M(r_2), M(r_3), M(r_4), M(r_5))$ должен характеризовать латентную переменную (удовлетворенность объяснением языковой модели). Но для этого необходимо, чтобы оценки группы респондентов были согласованы. Согласованность проверяется с помощью коэффициента α -Кронбаха [14]. Значение коэффициента меньше 0,5 означает, что оценки в группе не согласованы и вектор математических ожиданий не характеризует латентную переменную. Нужно отбросить противоречивые ответы и снова проверить оценки на согласованность, или собрать новую группу респондентов, или изменить вопросы, провести новый опрос и заново определить согласованность оценок. Эта процедура выполняется до получения приемлемого значения коэффициента согласованности.

Построим модель группы респондентов-единомышленников. Пусть имеем респондента R , который дал свою индивидуальную оценку $R = (r_1, r_2, r_3, r_4, r_5)$, $r_i \in \{1, 2, 3, 4, 5, 6, 7\}$, $i \in \{1, 2, 3, 4, 5\}$. Этого респондента будем называть базовым, а его оценку R базовой оценкой.

Используя оценки базового респондента, сформируем множество возможных оценок по следующему правилу: $R_j = (r_1 + x_{1j}, r_2 + x_{2j}, r_3 + x_{3j}, r_4 + x_{4j}, r_5 + x_{5j})$, где r_1, r_2, r_3, r_4, r_5 – оценки по критериям базового респондента; x_{ij} – случайная ошибка оценки j -го респондента по i -му критерию. Будем считать, что случайная ошибка респондента распределена по нормальному закону, т. е. $x_{ij} \in [-\sigma, +\sigma]$, где σ – среднеквадратическое отклонение ошибки оценивания.

Для эксперимента использовался генератор нормально распределенных случайных чисел из приложения Excel (функция НОРМ.ОБР(СЛЧИС()); $r_i; \sigma$), где r_i – оценка базового респондента по i -му критерию; σ – среднеквадратическое отклонение. В результате формировались 1000 гипотетических опросов, включающих 600 гипотетических респондентов со среднеквадратическими отклонениями ошибки оценивания $\sigma = 2$ и $\sigma = 3$. Для каждого случая (1000 опросов 600 респондентов для $\sigma = 2$, 1000 опросов 600 респондентов для $\sigma = 3$) был вычислен коэффициент согласованности α -Кронбаха (5) [14], который во всех случаях показывал неудовлетворительную согласованность опроса со значениями меньше 0,5 и даже с отрицательными значениями.

Далее из полученных гипотетических опросов выделялись только те анкеты, которые удовлетворяют неравенствам разумности респондента (1)–(4). При разных случайных генерациях оценок получалось от 40 до 70 (6–12 %) разумных анкет из 600 случайных анкет, построенных по определенному выше правилу. Эти анкеты образуют анкеты однородной группы разумных респондентов.

Для проверки согласованности оценок этой группы воспользуемся показателем α -Кронбаха [14]

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_x^2} \right), \quad (5)$$

где k – число вопросов, заданных респондентам; σ_i^2 – дисперсия оценок респондента по каждому вопросу; σ_x^2 – дисперсия суммы оценок респондентов.

Получилось, что среди 1000 опросов, из которых выделялись анкеты единомышленников базового респондента, у 97 % групп показатель согласованности α -Кронбаха превышает 0,7, что свидетельствует о хорошей согласованности.

Проверялся коэффициент α -Кронбаха для анкет, сгенерированных для двух значений среднеквадратического отклонения оценок респондентов от математического ожидания, $\sigma = 2$ и $\sigma = 3$. Результаты эксперимента показаны на рис. 2.

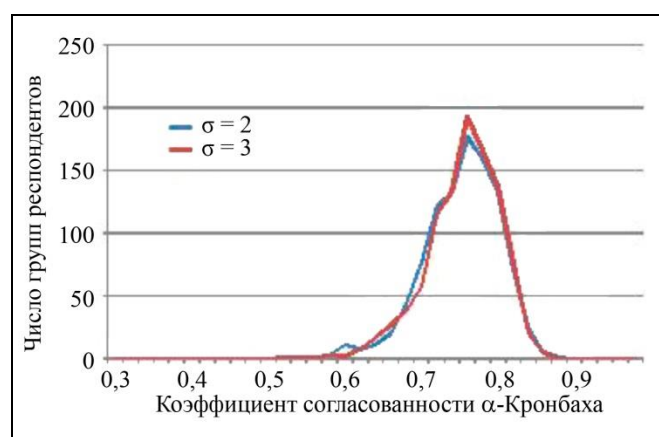


Рис. 2. Коэффициент α -Кронбаха для $\sigma = 2$ и $\sigma = 3$

Видно, что для разных значений среднеквадратического отклонения число согласованных опросов меняется незначительно, порядка 97 % опросов согласованы. Все это позволяет говорить об однородности групп респондентов, основанных на модели разумного респондента.

Для того чтобы множество используемых критериев позволяло измерить значение латентной переменной, нужно, чтобы они коррелировали между собой.

Определим ранговую корреляцию Спирмена для рассматриваемых критериев по формуле

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

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

Результат расчета ранговой корреляции (6) для 1000 опросов групп респондентов-единомышленников показан на рис. 3.

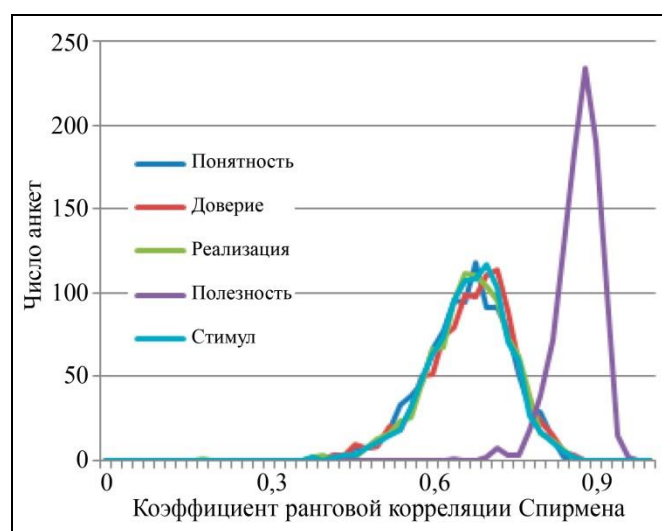


Рис. 3. Значение коэффициента ранговой корреляции Спирмена для всех критериев

Из графика видно, что 97 % анкет имеют ранговую корреляцию больше 0,5.

Для каждого коэффициента корреляции определим его статистическую значимость. По следующей формуле рассчитываем наблюдаемое значение t -критерия:

$$t_{obs} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \quad (7)$$

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

По таблице критических значений Стьюдента для значения n определяем критическое значение t_{cr} . Если наблюдаемое значение t -критерия (7) больше критического, $t_{obs} > t_{cr}$, то корреляция значима, т. е. гипотезу о существовании корреляции можно принять.

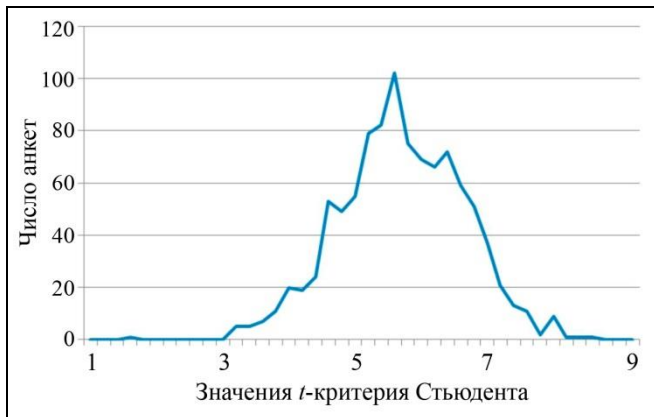


Рис. 4. Наблюдаемые значения t -критерия Стьюдента

Критические значения t -критерия для $n = 40$ и уровней значимости 0,05 и 0,01 соответственно равны $t_{cr(0,05)} = 2,02$ и $t_{cr(0,01)} = 3,55$, а для $n = 70$ соответственно $t_{cr(0,05)} = 2,0$ и $t_{cr(0,01)} = 3,46$. Судя по рис. 4, у 99 % анкет наблюдаемое значение t -критерия выше критических значений t -критерия. Следовательно, коэффициенты корреляции для всех оцениваемых критериев статистически значимые.

Необходимо отметить, что результаты реальных полевых исследований с подбором группы респондентов и их опросом проходят все вышеприведенные статистические исследования. При этом значения показателей согласованности, ранговой корреляции, статистической значимости, которые были получены для гипотетической группы, оцениваются как качественные. Отметим также, что в зависимости от цели опроса и предметной области модель разумного респондента может быть и другой. Однако если исследователь такую модель построил и обосновал, то она может быть полезна при обработке результатов полевых социологических или психологических исследований. Применение такой модели может работать в качестве своеобразного фильтра на этапах ремонта выборки (исключения некачественных анкет). Обычно при ремонте выборки эксперт, опираясь на собственную субъективную модель разумности респондента, исключает некорректные анкеты.

В этом случае отремонтированная выборка будет отражать субъективную модель разумности эксперта. Если эта субъективная модель формализована, то процесс ремонта выборки можно автоматизировать. Приведенные выше эксперименты показали, что из случайно сгенерированных оценок опроса можно получить согласованные анкеты с оценками латентной переменной, используя модель разумности респондента.

Таким образом, оценка объяснения языковой модели одним респондентом на основе модели его разумного поведения отражает поведение однородной группы респондентов-единомышленников и может быть использована на этапах выбора лучшей языковой модели как ассистента ЛППР.

4. ОЦЕНКА УДОВЛЕТВОРЕННОСТИ ЛИЦА, ПРИНИМАЮЩЕГО РЕШЕНИЕ, ОБЪЯСНЕНИЕМ ЯЗЫКОВОЙ МОДЕЛИ

Удовлетворенность объяснением с использованием шкалы Лайкерта оценивалась по пяти параметрам – это вектор оценок. Многокритериальная оценка может быть сведена к единственной числовой оценке. Методы многокритериального оценивания широко освещены в отечественной и зарубежной научной литературе. Например, самый простой – метод линейной свертки, суть которого заключается в определении коэффициентов важности для каждого критерия и нахождении суммы произведений оценки по критерию и коэффициента важности. Однако он обладает недостатками: в частности, из-за суммирования взвешенных оценок возникает эффект компенсации: важная, но низкая оценка, может быть компенсирована суммой неважных, но высоких оценок.

Поэтому определим экспертным способом две оценки, характеризующие идеальное объяснение для исследовательских целей и идеальное объяснение для практической цели [1]. Для экспертной оценки идеальных объяснений будем использовать шкалу Лайкерта, оценки представлены в табл. 2.

Таблица 2

Оценки идеальных объяснений для исследовательской и практической задач

Тип задачи	Понятность	Доверие	Реализация	Полезность	Стимул для любопытства
Исследовательская задача	Абсолютно понятно (7)	Скорее не доверяю (3)	Скорее не реализуется (3)	Очень полезно (7)	Обязательно возникает (7)
Прагматичная задача	Абсолютно понятно (7)	Абсолютно доверяю (7)	Легко реализуется (7)	Очень полезно (7)	Скорее не возникает (3)



Поясним эти оценки идеальных объяснений.

Для решения исследовательской задачи объяснение должно быть абсолютно понятно, очень полезно, обязательно возникает стимул для дальнейшего исследования, при этом достоверность и реализуемость не столь важны. Отметим, что в эвристических методах генерации решений контрфактические и нереализуемые альтернативы не отбрасываются, а рассматриваются. Считается, что они стимулируют выработку креативных решений и дают стимул для дальнейшего поиска. Поэтому достоверность и реализуемость для исследовательской задачи имеют низкие значения. Здесь происходит поиск объяснений, которые в конечном итоге должны привести к возможности обоснования и реализации самого решения.

Для решения прагматической задачи объяснение должно быть абсолютно понятно, очень полезно, абсолютно достоверно и реализуемо. При этом стимул для дальнейшего исследования должен уменьшаться (т. е. поиск закончен, приступаем к реализации).

Исследовательская задача решается до тех пор, пока не будут найдены легко реализуемые и абсолютно достоверные решения – в этом случае будем приближаться к идеальной точке для прагматической задачи.

Идеальное объяснение для решения исследовательской задачи – это вектор оценок $R_{inv} = (r_{inv1}, r_{inv2}, r_{inv3}, r_{inv4}, r_{inv5}) = (7, 3, 3, 7, 7)$. Идеальное объяснение для решения практической задачи – это вектор оценок $R_{prg} = (r_{prg1}, r_{prg2}, r_{prg3}, r_{prg4}, r_{prg5}) = (7, 7, 7, 7, 3)$.

Все возможные оценки по всем критериям – это все возможные векторы оценок из прямого произведения шкалы Лайкерта, $SR = \times_i E_i$, где $E_i = \{1, 2, 3, 4, 5\}$; i – номер критерия, $i \in \{1, 2, 3, 4, 5\}$. То есть $R_{prg}, R_{inv} \in SR$. Аналогично все оценки объяснений, данные j -м респондентом, $R_j = (r_{1j}, r_{2j}, r_{3j}, r_{4j}, r_{5j})$, $r_{ij} \in \{1, 2, 3, 4, 5, 6, 7\}$, принадлежат прямому произведению шкал Лайкерта, т. е. $R_j \in SR \forall j$.

Расстояние вектора оценки объяснения языковой модели R_j при решении исследовательской задачи и вектора оценки идеального объяснения для j -го вопроса R_{invj} определим как $St_{invj} = \rho(R_j, R_{invj})$, где ρ – метрика расстояния, $St_{invj} \in [0, \rho(R_0, R_{inv})]$.

Расстояние вектора оценки объяснения языковой модели R_j при решении практической задачи и вектора оценки идеального объяснения для j -го вопроса R_{prgj} определим как $St_{prgj} = \rho(R_j, R_{prgj})$, где ρ – метрика расстояния, $St_{prgj} \in [0, \rho(R_0, R_{prg})]$.

Здесь $R_0 \in SR$ вектор минимальных элементов множества SR , т. е. $R_0 = (1, 1, 1, 1, 1)$. В рассматри-

ваемом случае для исследовательской задачи $\rho(R_0, R_{inv}) = 22$, а для практической задачи $\rho(R_0, R_{prg}) = 26$.

В качестве метрики расстояния выбирается метрика «городских кварталов» Минковского. Тогда расстояние St_{invj} при решении исследовательской задачи определится по формуле

$$St_{invj} = \sum_{i=1}^5 |r_{ij} - r_{invj}|, \quad (8)$$

где j – номер вопроса; i – номер критерия оценивания.

Расстояние St_{prgj} при решении практической задачи определится по формуле

$$St_{prgj} = \sum_{i=1}^5 |r_{ij} - r_{prgj}|. \quad (9)$$

Удобно пользоваться оценкой достижения объяснения языковой модели идеального объяснения, выраженной в процентах.

Определение 1. Степенью $St_{invj}^{\%}$ удовлетворенности ЛПП объяснением языковой модели при решении исследовательской задачи, будем называть степень достижения вектором пользовательских оценок j -го объяснения идеального объяснения исследовательской задачи, выраженную в процентах, которая определяется из соотношения

$$St_{invj}^{\%} = \left(1 - \frac{St_{invj}}{\rho(R_0, R_{inv})}\right) \cdot 100\%. \quad (10)$$

Определение 2. Степенью $St_{prgj}^{\%}$ удовлетворенности ЛПП объяснением языковой модели при решении практической задачи, будем называть степень достижения вектором пользовательских оценок j -го объяснения идеального объяснения практической задачи, выраженную в процентах, которая определится из соотношения

$$St_{prgj}^{\%} = \left(1 - \frac{St_{prgj}}{\rho(R_0, R_{prg})}\right) \cdot 100\%. \quad (11)$$

С помощью выражений (10) и (11) можно вычислять и сравнивать степени удовлетворенности ЛПП по разным вопросам и для разных языковых моделей. Это позволит выбрать лучшую языковую модель, удовлетворяющую потребности ЛПП в процессах поддержки принятия решений.

5. ПРИМЕР

Для примера использования языковой модели в качестве ассистента ЛПП были рассмотрены две российские языковые модели: GigaChat 2.0 от Сбербанка и YandexGPT 5 Pro от Яндекса. Оба производителя заявляют о замечательных способностях их языковых моделей и предлагают как бесплатный вариант их использования, так и возможность приобретения лицензии на

дополнительные возможности языковой модели. Применялся бесплатный вариант языковых моделей, представленный на сайтах разработчиков [15, 16]. В эти модели включены системные промпты разработчиков языковых моделей. Пользовательские промпты не задействовались. Цель данного примера заключается в демонстрации метода оценки удовлетворенности пользователя объяснениями этих двух моделей.

Обеим моделям задавался один и тот же вопрос «почему?», из ответа модели GigaChat 2.0 формировался следующий вопрос «почему?» до тех пор, пока не появится понимание проблемы. Таким образом, реализуется метод выявления причин «пять почему?». Это исследовательская часть процесса принятия решений, которая формирует информационную среду поддержки принятия решений благодаря объяснениям языковой модели. Затем проводилась практическая часть, в которой ЛПР пытается получить дополнительные объяснения от языковой модели о возможности реализации рассматриваемой альтернативы. В этом случае языковой модели задаются вопросы «как?» или «какой?».

Языковые модели дают на вопросы развернутые ответы, которые вынесены в приложение. (Ответы моделей приведены в том виде, в котором они предоставлены автором, без редактирования и корректуры. – Прим. ред.). Далее приведены комментарии к ответам языковой модели, которые обосновывают экспертные оценки по пяти критериям.

Вопрос 1 «Почему у бизнесменов возникают финансовые проблемы?» Развернутые ответы двух языковых моделей приведены в табл. П1 в приложении.

Комментарий. Ответ языковой модели GigaChat 2.0 содержит краткую аннотацию ответа и следующие семь причин этой проблемы: недостаточный контроль расходов; нехватка капитала; несоблюдение налогового законодательства; плохое управление рисками; низкая рентабельность продукции или услуг; ошибочная маркетинговая политика; персонал и организация труда. Каждая из причин описана в некотором шаблоне: краткое

описание причины, возможные последствия и возможные действия для нейтрализации негативных последствий.

Оценки объяснения модели GigaChat 2.0 в ответ на первый вопрос приведены в табл. 3. Разработчики модели GigaChat 2.0 не раскрывают свои секреты, но здесь можно предположить, что применяется техника Few-shot-промптинга, встроенная в языковую модель.

В объяснении модели YandexGPT 5 Pro приведены две основные причины финансовых проблем: недостаточное планирование и анализ рынка; неэффективное управление финансами. Каждая из этих причин имеет объяснение; как правило, через запятую перечислены возможные последствия и трудности, если причины не будут устранены. Такое объяснение не очень понятно эксперту. Экспертная оценка для этой модели ниже, чем для модели GigaChat 2.0, и также представлена в табл. 3.

Оценки GigaChat 2.0 по вопросу 1 $R_1^{GigaChat} = (7, 6, 5, 5, 6)$.

Оценки YandexGPT 5 Pro по вопросу 1 $R_1^{GPT 5 Pro} = (6, 5, 5, 5, 7)$.

Вопрос 2 – практический, «как?» – был сформирован из описания причин финансовых проблем ответа модели GigaChat 2.0. Вопрос такой: «Как повысить эффективность контроля над расходами?»

Развернутые объяснения двух языковых моделей приведены в табл. П2 в приложении.

Комментарий. В объяснении модели GigaChat 2.0 приводятся два метода повышения контроля над расходами: бюджетирование; автоматизация учета и анализа расходов. По каждому методу объяснение представлено в виде шаблона, включающего описание метода и его преимущества. Это способствует пониманию объяснения. Такое объяснение выглядит правильным, но несколько абстрактным, что скорее способствует разработке стратегии развития бизнеса. Поэтому несколько снижена оценка по критерию реализуемости объяснения. Оценки приведены в табл. 4.

Таблица 3

Оценки объяснений языковых моделей по вопросу 1

Модель	Понятность	Доверие	Реализация	Полезность	Стимул для любопытства
GigaChat 2.0	Абсолютно понятно (7)	Доверяю (6)	Трудно реализуется (5)	Малополезно (5)	Возникает (6)
YandexGPT 5 Pro	Понятно (6)	Слабо доверяю (5)	Трудно реализуется (5)	Малополезно (5)	Обязательно возникает (7)

Таблица 4

Оценки объяснений языковых моделей по вопросу 2

Модель	Понятность	Доверие	Реализация	Полезность	Стимул для любопытства
GigaChat 2.0	Абсолютно понятно (7)	Доверяю (6)	Реализуется (6)	Полезно (6)	Возникает (6)
YandexGPT 5 Pro	Понятно (6)	Слабо доверяю (5)	Легко реализуется (7)	Полезно (6)	Возникает (6)



В объяснении модели YandexGPT 5 Pro также содержится описание двух методов решения финансовых проблем: внедрение системы бюджетирования и финансового планирования; анализ и оптимизация структуры расходов. Для каждого метода приводятся описание действий, направленных на реализацию этих методов. Поэтому такое объяснение больше подходит для принятия решений по оперативному управлению бизнесом. Здесь оценки реализуемости объяснения высокие, они также приведены в табл. 4.

Оценки GigaChat 2.0 по вопросу 2 $R_2^{GigaChat} = (7, 6, 6, 6, 6)$.

Оценки YandexGPT 5 Pro по вопросу 2 $R_2^{GPT 5 Pro} = (6, 5, 7, 6, 6)$.

Вопрос 3 языковым моделям – практический и направлен на обоснование альтернативы «бюджетирование» для решения финансовой проблемы в бизнесе. Вопрос следующий: «Как внедрить бюджетирование на малом предприятии? Укажи два метода.» Развернутые ответы приведены в табл. ПЗ в приложении.

Комментарий. В объяснении модели GigaChat 2.0 приводятся два метода внедрения: сверху вниз и снизу вверх. Традиционно для этой языковой модели объяснение представляется в шаблоне, включающем описание метода, его преимущества и недостатки. Объяснение понятно, полезно, оценка по критерию «реализация» несколько занижена. Оценки приведены в табл. 5.

В объяснении модели YandexGPT 5 Pro также содержится описание двух методов: создание простого и понятного бюджета с основными статьями доходов и расходов и назначение лица, ответственного за бюджетирование и контроль его исполнения. Приводится объяснение каждого метода. Уже по названиям самих методов в этой языковой модели видна их практическая направленность. То есть объяснения этих методов со-

держат практические советы, позволяющие реализовать бюджетирование в бизнесе. Оценки объяснения этой модели также приведены в табл. 5.

Оценки GigaChat 2.0 по вопросу 3 $R_3^{GigaChat} = (7, 7, 6, 6, 5)$.

Оценки YandexGPT 5 Pro по вопросу 3 $R_3^{GPT 5 Pro} = (6, 6, 7, 6, 5)$.

Вопрос 4 языковым моделям также прагматический и направлен на получение объяснения о цене внедрения бюджетирования. Вопрос следующий: «Какова цена внедрения бюджетирования на малом предприятии?» Развернутый ответ приведен в табл. П4 в приложении.

Комментарий. В объяснении модели GigaChat 2.0 приводится структура цены внедрения бюджетирования в бизнесе. Это основные затраты на программное обеспечение, консалтинговые услуги, обучение персонала, подготовку регламентов и внутренних документов, общую оценку затрат. По каждой статье затрат приводятся объяснение и рекомендации по выбору программных продуктов и их стоимости, стоимости консультаций и обучения персонала и др. Приводятся границы цены внедрения бюджетирования. Конечно, все рекомендации и цены требуют проверки, но в целом оценки объяснения по этому критерию высокие, они приведены в табл. 6.

В объяснении модели YandexGPT 5 Pro содержатся советы и рекомендации по определению стоимости внедрения бюджетирования в бизнесе в зависимости от размера бизнеса и квалификации персонала. Объяснение включает возможные статьи затрат без указания их стоимости. Оценки объяснения этой модели также представлены в табл. 6.

Оценки GigaChat 2.0 по вопросу 4 $R_4^{GigaChat} = (7, 5, 6, 6, 6)$.

Оценки YandexGPT 5 Pro по вопросу 4 $R_4^{GPT 5 Pro} = (6, 5, 7, 5, 6)$. ♦

Таблица 5

Оценки объяснений языковых моделей по вопросу 3

Модель	Понятность	Доверие	Реализация	Полезность	Стимул для любопытства
GigaChat 2.0	Абсолютно понятно (7)	Абсолютно доверяю (7)	Реализуется (6)	Полезно (6)	Слабо возникает (5)
YandexGPT 5 Pro	Понятно (6)	Доверяю (6)	Легко реализуется (7)	Полезно (6)	Слабо возникает (5)

Таблица 6

Оценки объяснений языковых моделей по вопросу 4

Модель	Понятность	Доверие	Реализация	Полезность	Стимул для любопытства
GigaChat 2.0	Абсолютно понятно (7)	Слабо доверяю (5)	Реализуется (6)	Полезно (6)	Возникает (6)
YandexGPT 5 Pro	Понятно (6)	Слабо доверяю (5)	Легко реализуется (7)	Малополезно (5)	Возникает (8)

6. АНАЛИЗ ОЦЕНОК УДОВЛЕТВОРЕННОСТИ ОБЪЯСНЕНИЯМИ ЯЗЫКОВЫХ МОДЕЛЕЙ GIGACHAT 2.0 И YANDEXGPT 5 PRO

Экспертные оценки качества объяснения языковых моделей GigaChat 2.0 по четырем вопросам приводятся в табл. 7. Здесь в строках для каждого вопроса приведены экспертные оценки по выделенным в § 1 критериям. В трех последних столбцах указаны: значения расстояния St_{invj} и через дробную черту степень удовлетворенности эксперта $St_{invj}^{\%}$ объяснением языковой модели при решении исследовательской задачи; значения расстояния St_{prgj} и через дробную черту степень удовлетворенности эксперта $St_{prgj}^{\%}$ объяснением языковой модели при решении практической задачи; разность $St_{prgj}^{\%} - St_{invj}^{\%}$, позволяющая оценить качество удовлетворенности объяснением при решении исследовательской и практической задач.

Оценки одного респондента объяснений языковой модели GigaChat 2.0 соответствуют модели разумного поведения. В табл. 7 приведены значения коэффициента согласованности α -Кронбаха для ответов на каждый вопрос для гипотетической группы единомышленников и коэффициенты ранговой корреляции для каждого критерия.

Как видно из таблицы объяснения языковой модели, модель GigaChat 2.0 ближе к идеальному объяснению практической задачи, чем к идеальному объяснению исследовательской задачи. Справедливости ради отметим, что только первый вопрос «почему?» относится к решению исследовательской задачи, остальные вопросы – «как?» и «какой?» – относятся к решению практической задачи. Как видно из табл. 7, удовлетворенность объяснениями по вопросам 2 и 3 при решении практической задачи растет с 76,9 % до 84,6 %.

Это может говорить о расширении информационной среды эксперта для принятия решений. Однако объяснение в ответ на вопрос 4 снизило удовлетворенность до 71,1 %. В исследовательском процессе поиска решений это нормальное явление.

Оценки одного респондента объяснений языковой модели YandexGPT 5 Pro соответствуют модели разумного поведения. В табл. 8 приведены значения коэффициента согласованности α -Кронбаха для ответов на каждый вопрос для гипотетической группы единомышленников и коэффициенты ранговой корреляции для каждого критерия.

В строках табл. 8, так же, как и в табл. 7, приведены экспертные оценки по выделенным в § 1 критериям; значения расстояния St_{invj} и степень удовлетворенности эксперта $St_{invj}^{\%}$ объяснением языковой модели при решении исследовательской задачи; значения расстояния St_{prgj} и степень удовлетворенности эксперта $St_{prgj}^{\%}$ объяснением языковой модели при решении практической задачи; разность $St_{prgj}^{\%} - St_{invj}^{\%}$, позволяющая оценить качество удовлетворенности объяснением при решении исследовательской и практической задач.

Как видно из табл. 8, объяснения языковой модели YandexGPT 5 Pro ближе к идеальному объяснению практической задачи, чем к идеальному объяснению исследовательской задачи, за исключением первого вопроса. Удовлетворенность объяснениями по вопросам 2 и 3 при решении практической задачи растет от 73,1 % до 80,8 %. Это говорит о расширении информационной среды эксперта для принятия решений, но объяснение по вопросу 4 снизило удовлетворенность до 69,2 %.

Теперь сравним степени удовлетворенности эксперта объяснениями языковых моделей GigaChat 2.0 и YandexGPT 5 Pro отдельно по удовлетворенности при решении исследовательской и практической задач. Результаты сравнения представлены в табл. 9.

Таблица 7

Оценки по критериям и степень удовлетворенности объяснениями языковой модели GigaChat 2.0

Вопрос / коэф. α - Кронбаха	Понятность / коэф. корр.	Доверие / коэф. корр.	Реализа- ция / ко- эф. корр.	Полезность / коэф. корр.	Стимул для любопытства / коэф. корр.	Иссл. задача. расст./ удовл. $St_{invj}^{\%}$	Практ. задача. расст./ удовл. $St_{prgj}^{\%}$	$St_{prgj}^{\%} -$ $St_{invj}^{\%}$
Вопрос 1 $\alpha = 0,80$	7 $\rho = 0,62$	6 $\rho = 0,80$	5 $\rho = 0,77$	5 $\rho = 0,86$	6 $\rho = 0,67$	8 63,6 %	8 69,2 %	5,6 %
Вопрос 2 $\alpha = 0,82$	7 $\rho = 0,76$	6 $\rho = 0,68$	6 $\rho = 0,67$	6 $\rho = 0,90$	6 $\rho = 0,67$	8 63,6 %	6 76,9 %	13,3 %
Вопрос 3 $\alpha = 0,81$	7 $\rho = 0,75$	7 $\rho = 0,77$	6 $\rho = 0,65$	6 $\rho = 0,90$	5 $\rho = 0,76$	10 54,5 %	4 84,6 %	30,1 %
Вопрос 4 $\alpha = 0,79$	7 $\rho = 0,73$	5 $\rho = 0,70$	6 $\rho = 0,83$	6 $\rho = 0,84$	6 $\rho = 0,70$	7 62,2 %	7 71,1 %	4,9 %



Таблица 8

Оценки по критериям и степень удовлетворенности объяснениями языковой модели YandexGPT 5 Pro

Вопрос / коэф. α - Кронбаха	Понятность / коэф. корр.	Доверие / коэф. корр.	Реализа- ция / ко- эф. корр.	Полезность / коэф. корр.	Стимул для любопытства / коэф. корр.	Иссл. задача. Расст. / удовл. $St_{invj}^{\%}$	Практ. задача. Расст. / удовл. $St_{prgj}^{\%}$	$St_{prgj}^{\%} -$ $St_{invj}^{\%}$
Вопрос 1 $\alpha = 0,71$	6 $\rho = 0,56$	5 $\rho = 0,73$	5 $\rho = 0,57$	5 $\rho = 0,87$	7 $\rho = 0,60$	7 68,2 %	11 57,7 %	-10,5 %
Вопрос 2 $\alpha = 0,80$	6 $\rho = 0,65$	5 $\rho = 0,73$	7 $\rho = 0,66$	6 $\rho = 0,87$	6 $\rho = 0,73$	9 59,1 %	7 73,1 %	14,0 %
Вопрос 3 $\alpha = 0,82$	6 $\rho = 0,67$	6 $\rho = 0,70$	7 $\rho = 0,79$	6 $\rho = 0,80$	5 $\rho = 0,71$	11 50,0 %	5 80,8 %	30,8 %
Вопрос 4 $\alpha = 0,71$	6 $\rho = 0,58$	5 $\rho = 0,74$	7 $\rho = 0,59$	5 $\rho = 0,85$	6 $\rho = 0,61$	10 54,5 %	8 69,2 %	14,7 %

Таблица 9

Сравнение степени удовлетворенности эксперта объяснениями языковых моделей при решении исследовательской и практической задач

Вопрос	Исследовательская задача			Практическая задача		
	$St_{invj_GigaChat}^{\%}$	$St_{invj_GPT\ 5\ Pro}^{\%}$	$St_{invj_GigaChat}^{\%} -$ $St_{invj_GPT\ 5\ Pro}^{\%}$	$St_{prgj_GigaChat}^{\%}$	$St_{prgj_GPT\ 5\ Pro}^{\%}$	$St_{prgj_GigaChat}^{\%} -$ $St_{prgj_GPT\ 5\ Pro}^{\%}$
1	63,6	68,2	-4,5	69,2	57,7	11,5
2	63,6	59,1	4,5	76,9	73,1	3,8
3	54,5	50,0	4,5	84,6	80,8	3,8
4	68,2	54,5	13,6	73,1	69,2	3,8

Здесь в строках приведены номера вопросов и степени удовлетворенности эксперта объяснениями языковых моделей GigaChat 2.0 и YandexGPT 5 Pro при решении исследовательской или практической задачи. Разность $St_{invj_GigaChat}^{\%} - St_{invj_GPT\ 5\ pro}^{\%}$ в третьем столбце показывает отличие в степени удовлетворенности эксперта объяснениями при решении исследовательской задачи. По вопросам 2, 3 и 4 преимущество имеет языковая модель GigaChat 2.0. Разность $St_{invj_GigaChat}^{\%} - St_{invj_GPT\ 5\ pro}^{\%}$ в шестом столбце показывает отличие в степени удовлетворенности эксперта объяснениями при решении практической задачи. По вопросам 1, 2, 3 и 4 небольшое преимущество также имеет языковая модель GigaChat 2.0.

Из табл. 9 видно, что по вопросу 1 «Почему у бизнесменов возникают финансовые проблемы?» с точки зрения достижения исследовательской цели объяснение модели YandexGPT 5 Pro имеет некоторое преимущество (68,2 %) по отношению к объяснению модели GigaChat 2.0 (63,6 %). В тоже время для достижения практической цели объяснение по этому же вопросу модели GigaChat 2.0 (69,2 %) оказалось лучше объяснения модели Yan-

dexGPT 5 Pro (57,7 %). На вопрос 4 «Какова цена внедрения бюджетирования на малом предприятии?» объяснение для достижения исследовательской цели поддержки принятия решений модели GigaChat 2.0 (68,2 %) оказалось лучше объяснения YandexGPT 5 Pro (54,5 %). Оценка объяснений по вопросам 2 и 3 для исследовательской и практической задач незначительно отличаются в пользу модели GigaChat 2.0.

Необходимо отметить, что приведенные результаты анализа и оценки отражают качество сложной когнитивной системы, включающей ЛПР и языковую модель. Эти оценки зависят не только от способностей языковой модели, но и от уровня знаний самого эксперта. Другой эксперт или группа экспертов могут получить другие оценки и другие предпочтения в выборе языковой модели в качестве ассистента в поддержке принятия решений.

7. ОБСУЖДЕНИЕ

В первой части статьи [1] были выделены классы моделей объяснений: дедуктивно-номологическая модель объяснения, унификационистские,

прагматические. Они исследованы философами; конечно, общей модели научного объяснения у них нет, споры не утихают. Но понятны основные свойства этих классов и их полезность для поддержки принятия решения, рассмотренная в первой части статьи. Языковая модель генерирует текст объяснения. Можно считать этот текст объяснением с точки зрения методологии науки в контексте объяснений явлений или нет?

Проанализируем связь объяснений, полученных языковыми моделями, с моделями объяснения, предложенными в методологии науки.

Объяснения языковой модели GigaChat 2.0 хорошо структурированы, вписаны в некоторый шаблон. Они понятны и полны, но абстрактны. В методологии науки есть унификационистские модели объяснения [1, 17], которые строят объяснения по шаблону. Эти модели не содержат логического вывода и поэтому проверить достоверность объяснений этих моделей трудно. Однако объяснение в унификационистской модели объяснения выводится на основе объяснительного ресурса, который содержит проверенные и достоверные сведения. Если в качестве объяснительного ресурса языковая модель GigaChat 2.0 использует нейронную сеть, обученную текстами справочников и энциклопедий, тогда доверие к объяснениям языковой модели GigaChat 2.0 будет высоким и полезным для расширения информационной среды ЛПР.

Объяснения языковой модели YandexGPT 5 Pro структурированы, понятны. Здесь объяснения имеют практическую направленность, как будто они обучены опытом хороших и успешных менеджеров или консультантов. В методологии науки есть прагматические модели объяснения [18], они описаны в первой части статьи [1] и основаны на опыте людей и эмпирических наблюдаемых данных. Объяснения таких моделей характеризуются эмпирической адекватностью [18]. Принятие таких объяснений основано на вере, уверенности, что они верны. Языковую модель YandexGPT 5 Pro можно ассоциировать с моделью, генерирующей прагматические объяснения. Доверие к таким объяснениям будет ниже. Нужно учитывать возможные предубеждения экспертов, знаниями которых была обучена или дообучена модель.

Как видно, экспертным способом эти две языковые модели можно отнести к разным классам моделей объяснения из методологии науки. Принадлежность объяснений языковой модели позволяет в самом общем виде судить о достоверности объяснений, проблемах и перспективах их развития.

В первой части этой статьи была рассмотрена дедуктивно-номологическая модель объяснения [1, 19]. Эта классическая модель научного объяснения основана на дедуктивном выводе, базирующемся на известных законах. Судя по всему, в приведенных выше объяснениях строгого логического вывода на основании законов не обнаружено. В качестве примера исследуемым языковым моделям был задан вопрос на знание экономических законов: «По какому экономическому закону у предприятия возникают финансовые проблемы?» Объяснения моделей приведены в табл. П5 в приложении.

Как видно, обе модели ссылаются на один и тот же экономический закон – это закон соответствия доходов расходам – и приводят основные причины его нарушения. Однако строгого логического вывода здесь нет. Такой вывод способны дать рассуждающие языковые модели, которые здесь не рассмотрены.

ЗАКЛЮЧЕНИЕ

Рассмотрены вопросы измерения качества гибридной системы поддержки принятия решений, включающей ЛПР и в качестве ассистента большую языковую модель. Предложены множество критериев оценки качества гибридной системы поддержки принятия решений и экспертная процедура оценивания. Особенностью экспертной оценки качества системы является предложенная структурная модель разумного эксперта. Показано, что модель разумного респондента позволяет создать виртуальную группу респондентов, обладающую свойствами согласованности и однородности. Показано, что оценки, основанные на модели разумного респондента, могут быть использованы в качестве оценок качества объяснения языковой модели. Формально определена степень удовлетворенности эксперта объяснением языковой модели при решении исследовательской и практической задач.

Приводится пример работы и оценивания системы поддержки принятия решений, включающей эксперта и языковые модели GigaChat 2.0 и YandexGPT 5 Pro.

Анализ объяснений этих языковых моделей в ответ на поставленные экспертом вопросы показывает, что эти объяснения можно отнести к моделям объяснения унификационистского (GigaChat 2.0) и прагматического типа (YandexGPT 5 Pro). Объяснения унификационистской модели базируются на так называемом объяснительном резерве, в каче-



стве которого могут выступать проверенные сведения из справочников и энциклопедий. Таким объяснениям можно доверять, однако они носят стратегический характер, их трудно адаптировать для решения оперативных задач. Объяснения прагматической модели базируются на опыте менеджеров, экспертов и эмпирических данных. Они полезны для решения оперативных задач управления, однако могут включать экспертные заблуждения и приводить к ошибкам. Объяснения этих двух моделей больше направлены на расширение информационной среды принятия решений и генерацию альтернатив.

Для обоснования альтернатив решений подходит дедуктивно-номологическая модель объяснений, основанная на логическом выводе и известных законах. Рассмотренные языковые модели не представляли логического обоснования альтернатив решений. Для получения дедуктивно-номологической модели объяснения необходим рассуждающий класс языковых моделей, а также специальные техники промптинга.

Дальнейшие исследования систем поддержки принятия решений, включающих ЛПР и большую языковую модель, видятся следующим образом:

- разработка промптов для поддержки объяснений, решающих исследовательскую задачу с применением эвристических методов структуризации информации, стратегического планирования, развития критического мышления и креативности, таких как SWOT-анализ, диаграммы Исикавы (*fishbone*) и др.;

- разработка промптов для реализации дедуктивно-номологической модели объяснения.

Предложенный метод измерения качества и оценки удовлетворенности ЛПР, основанный на структурной модели разумного эксперта и квази-эксперимента оценивания, может быть полезен на этапе создания системы поддержки принятия решений, включающей большую языковую модель. Этот метод позволит дать качественные оценки гибридным системам поддержки принятия решений и сократить время и стоимость оценивания без привлечения групп экспертов.

ПРИЛОЖЕНИЕ

Таблица III

Объяснения моделей в ответ на вопрос «Почему у бизнесменов возникают финансовые проблемы?»

GigaChat 2.0	YandexGPT 5 Pro
<p>Внутренние факторы</p> <p>1. Недостаточный контроль расходов Многие предприниматели сталкиваются с проблемами из-за отсутствия эффективного контроля над расходами. Это включает чрезмерные траты на ненужные активы, неэффективность операций и недостаток анализа финансовой отчетности. Важно регулярно анализировать расходы и искать возможности для оптимизации затрат.</p> <p>2. Нехватка капитала Недостаточное финансирование бизнеса также является частой причиной финансовых проблем. Отсутствие достаточного стартового капитала или невозможность привлечь дополнительные инвестиции может привести к нехватке денежных средств для покрытия текущих обязательств и расширения бизнеса.</p> <p>3. Несоблюдение налогового законодательства Налоговая дисциплина важна для любого предпринимателя. Нарушение налоговых правил или неправильное ведение бухгалтерии может привести к штрафам и дополнительным финансовым обязательствам, что усугубляет финансовую ситуацию.</p> <p>4. Плохое управление рисками Отсутствие стратегии управления рисками может привести к значительным потерям. Например, непредвиденные изменения на рынке, колебания курсов валют или потеря ключевых клиентов могут негативно сказаться на финансовом положении предприятия.</p> <p>5. Низкая рентабельность продукции или услуг</p>	<p>У бизнесменов могут возникать финансовые проблемы по разным причинам. Вот две из наиболее распространенных.</p> <ul style="list-style-type: none"> • Недостаточное планирование и анализ рынка. Отсутствие детального бизнес-плана, включающего анализ рынка, конкурентов, целевой аудитории и финансовых прогнозов, может привести к неправильным решениям и неэффективному использованию ресурсов. Неучтенные факторы, такие как изменения в законодательстве, экономические колебания или появление новых конкурентов, могут негативно сказаться на финансовом состоянии бизнеса. • Неэффективное управление финансами. Проблемы с управлением денежными потоками, бюджетом и инвестициями могут привести к нехватке средств для покрытия расходов, задолженности перед кредиторами и другими финансовыми трудностями. Неумение правильно распределять ресурсы, оптимизировать затраты и контролировать финансы может стать причиной серьезных финансовых проблем

<p>Если бизнес предлагает продукцию или услуги с низкой маржинальностью, это может стать источником постоянных финансовых проблем. Предприниматели должны внимательно следить за своей продуктовой линейкой и своевременно реагировать на изменения спроса и конкуренции.</p> <p>6. Ошибочная маркетинговая политика</p> <p>Неправильная оценка рынка и потребительских предпочтений может привести к убыткам. Эффективная маркетинговая стратегия должна учитывать потребности целевой аудитории и адаптироваться к изменениям рыночной среды.</p> <p>7. Персонал и организация труда</p> <p>Проблемы с персоналом, такие как низкая производительность сотрудников, высокая текучка кадров или недостаточная квалификация работников, могут существенно повлиять на прибыльность бизнеса. Грамотное управление человеческими ресурсами помогает минимизировать эти риски.</p> <p>Эти внутренние факторы являются ключевыми причинами возникновения финансовых проблем у многих российских предпринимателей. Для предотвращения таких ситуаций важно развивать компетенции в области финансового менеджмента, стратегического планирования и риск-менеджмента</p>	
--	--

Таблица П2

Объяснения моделей в ответ на вопрос «Как повысить эффективность контроля над расходами?»

GigaChat 2.0	YandexGPT 5 Pro
<p>Для повышения эффективности контроля над расходами можно применить следующие методы.</p> <p>Метод 1: Бюджетирование</p> <p>Описание: Регулярная разработка детального бюджета позволяет заранее планировать расходы и отслеживать фактические затраты относительно запланированных показателей. Это помогает выявлять отклонения и своевременно реагировать на них.</p> <p>Преимущества:</p> <ul style="list-style-type: none"> – позволяет установить лимиты расходов по различным статьям; – обеспечивает прозрачность финансовых потоков; – помогает избежать перерасхода средств. <p>Метод 2: Автоматизация учета и анализа расходов</p> <p>Описание: использование специализированных программ и сервисов для автоматизации процесса учета расходов упрощает сбор и обработку финансовой информации; современные системы позволяют автоматически классифицировать транзакции, формировать отчеты и анализировать тенденции.</p> <p>Преимущества:</p> <ul style="list-style-type: none"> – сокращение ошибок ручного ввода данных; – повышение скорости обработки информации; – возможность оперативного принятия решений на основе аналитической отчетности; – эти методы способствуют повышению точности и оперативности контроля над расходами, обеспечивая финансовую устойчивость организации 	<p>Внедрение системы бюджетирования и финансового планирования.</p> <ul style="list-style-type: none"> • Разработка детального бюджета на определенный период (месяц, квартал, год) с учетом всех предполагаемых доходов и расходов. • Регулярный мониторинг фактических расходов и сравнение их с бюджетными показателями для выявления отклонений и принятия корректирующих мер. • Использование программного обеспечения для автоматизации процесса бюджетирования и упрощения анализа финансовых данных. <p>Анализ и оптимизация структуры расходов:</p> <ul style="list-style-type: none"> • Проведение регулярного анализа всех статей расходов для выявления возможностей оптимизации. Это может включать пересмотр контрактов с поставщиками, поиск более выгодных условий закупки товаров и услуг, а также исключение ненужных или малоэффективных расходов. <p>Внедрение принципов бережливого производства и управления затратами, направленных на сокращение издержек без ущерба для качества продукции или услуг. Например, можно оптимизировать процессы, сократить потери и повысить эффективность использования ресурсов</p>



**Объяснения моделей в ответ на вопрос
«Как внедрить бюджетирование на малом предприятии? Укажи два метода»**

GigaChat 2.0	YandexGPT 5 Pro
<p>Для внедрения бюджетирования на малом предприятии существуют различные методы, среди которых наиболее распространенными являются методы «сверху-вниз» и «снизу-вверх».</p> <p>Метод «сверху-вниз»</p> <p>При данном методе руководство предприятия определяет общие цели и направления развития бизнеса, устанавливает ключевые показатели эффективности (KPI), распределяет ресурсы между подразделениями и сотрудниками. После утверждения бюджетов верхнего уровня, подразделения разрабатывают собственные детализированные планы исходя из поставленных целей и выделенных ресурсов.</p> <p>Преимущества метода:</p> <ul style="list-style-type: none"> • быстрое внедрение, • четкое распределение ответственности, • возможность учитывать стратегические приоритеты руководства. <p>Недостатки:</p> <ul style="list-style-type: none"> • недостаточная вовлеченность сотрудников нижних уровней управления, • возможная нехватка деталей и реалистичности бюджета. <p>Метод «снизу-вверх»</p> <p>Этот метод предполагает, что сначала сотрудники подразделений формируют собственные бюджеты, основываясь на текущих показателях своей деятельности, планах продаж, затрат и иных параметрах. Затем полученные данные обобщаются руководством и согласовываются на верхнем уровне управления предприятием.</p> <p>Преимущества метода:</p> <ul style="list-style-type: none"> • повышение точности планирования благодаря детальному знанию ситуации в каждом подразделении, • активизация вовлеченности персонала в процесс бюджетирования. <p>Недостатки:</p> <ul style="list-style-type: none"> • более длительный период согласования и принятия решений, • риск появления противоречий и несогласованности между различными подразделениями. <p>Таким образом, выбор подходящего метода зависит от специфики вашего малого предприятия, готовности персонала и требований руководства. Рекомендуется комбинировать оба подхода – устанавливать верхнеуровневые ориентиры сверху, одновременно стимулируя участие нижестоящих сотрудников в разработке конкретных планов</p>	<p>Создание простого и понятного бюджета с основными статьями доходов и расходов.</p> <ul style="list-style-type: none"> • Определите ключевые статьи доходов (например, продажи товаров или услуг) и расходов (аренда, зарплаты, закупки материалов, коммунальные услуги и т. д.). • Используйте электронные таблицы (например, Excel) для составления бюджета, так как они позволяют легко вносить изменения и проводить анализ данных. <p>Начните с составления бюджета на короткий период (например, месяц или квартал), чтобы иметь возможность быстро корректировать его в зависимости от изменений в бизнесе.</p> <p>Назначение ответственного лица за бюджетирование и контроль над его исполнением.</p> <ul style="list-style-type: none"> • Назначьте сотрудника или группу сотрудников, которые будут отвечать за составление бюджета, его мониторинг и анализ отклонений. • Обеспечьте обучение ответственного лица основам бюджетирования и финансового анализа, чтобы он мог эффективно выполнять свои обязанности. • Регулярно проводите встречи для обсуждения текущего состояния бюджета, анализа отклонений и планирования корректирующих действий. Это поможет поддерживать контроль над финансовыми потоками и своевременно реагировать на изменения

Объяснения моделей в ответ на вопрос «Какова цена внедрения бюджетирования на малом предприятии?»

GigaChat 2.0	YandexGPT 5 Pro
<p>Цена внедрения системы бюджетирования на малом предприятии зависит от нескольких факторов.</p> <p>Основные затраты</p> <p>1. Программное обеспечение</p> <p>Выбор программного продукта играет ключевую роль в формировании стоимости проекта. Вот некоторые варианты ПО:</p> <ul style="list-style-type: none">• бесплатные решения: есть бесплатные инструменты, такие как EXcel или OpenOffice Calc, однако их функциональность ограничена и требует высокой квалификации сотрудников;• платные продукты начального уровня: например, отечественный продукт «ФинГрад», цены которого начинаются примерно от 15–30 тыс. руб. в зависимости от функционала и количества пользователей;• специализированные CRM/ERP системы с модулем бюджетирования: эти системы часто включают дополнительные модули учета и аналитики, но стоимость начинается от 80–150 тыс. руб., иногда дороже – около 200–300 тыс. руб. Примеры российских решений: «1С Управление нашей фирмой», «Битрикс24». <p>2. Консалтинговые услуги</p> <p>Для правильной настройки и адаптации программы часто привлекают консультантов, аналитиков или аудиторов. Стоимость услуг варьируется значительно:</p> <ul style="list-style-type: none">• средняя ставка специалиста составляет порядка 2–5 тыс. руб./час;• проект может занять от пары недель до месяца или даже дольше, в зависимости от сложности предприятия;• итоговая сумма расходов на консалтинговую поддержку составит от 100 до 500 тыс. руб. и выше. <p>3. Обучение персонала</p> <p>Чтобы сотрудники эффективно использовали систему, важно провести обучение:</p> <ul style="list-style-type: none">• внутреннее обучение силами специалистов фирмы: бесплатно или минимальные расходы;• внешнее корпоративное обучение: обычно обходится в сумму от 10 до 50 тыс. руб. за группу участников. <p>4. Подготовка регламентов и внутренних документов</p> <p>Процесс подготовки регламентирующих процедур также несет расходы:</p> <ul style="list-style-type: none">• если самостоятельно разрабатываются процессы, эта работа бесплатна;• привлечение внешних экспертов обойдется дополнительно в 50–150 тыс. руб. <p>Дополнительные факторы влияния на цену</p> <ul style="list-style-type: none">• размер бизнеса: чем меньше предприятие, тем ниже требования к функциональности и возможностям ПО, соответственно дешевле внедрение;• автоматизация существующих процессов: уровень автоматизации учетных операций влияет на сложность интеграции бюджетирования;• ресурсная база: наличие квалифицированных кадров снижает потребность в привлечении сторонних специалистов.	<p>Цена внедрения бюджетирования на малом предприятии может варьироваться в зависимости от ряда факторов.</p> <ul style="list-style-type: none">• Размер предприятия и сложность бизнес-процессов: чем больше масштабы деятельности и сложнее структура бизнеса, тем больше ресурсов потребуется для внедрения бюджетирования.• Наличие программного обеспечения: использование специализированного программного обеспечения для бюджетирования может потребовать дополнительных затрат на покупку лицензий или разработку индивидуальной системы. Однако это может упростить процесс и повысить его эффективность.• Обучение персонала: затраты на обучение сотрудников, которые будут заниматься бюджетированием, также могут входить в общую стоимость. Это могут быть внутренние тренинги или привлечение внешних специалистов.• Консультационные услуги: в некоторых случаях предприятия могут привлекать внешних консультантов для помощи в разработке и внедрении системы бюджетирования, что также добавит к общим затратам.• Время и ресурсы руководства и сотрудников: внедрение бюджетирования требует времени и внимания со стороны руководства и ключевых сотрудников. Это внутренние ресурсы, которые также следует учитывать. <p>В целом, точные цифры затрат зависят от конкретных условий и потребностей предприятия. Для некоторых малых предприятий внедрение бюджетирования может обойтись минимальными затратами, особенно если использовать простые инструменты и существующие ресурсы, в то время как для других затраты могут быть значительными из-за необходимости приобретения специализированного ПО и привлечения дополнительных специалистов.</p>



Окончание табл. П4

<p>Общая оценка затрат</p> <p>Минимальная стоимость реализации базового проекта бюджетирования на небольшом предприятии может составить:</p> <ul style="list-style-type: none"> • от 70–100 тыс. руб. (самостоятельная настройка бесплатных инструментов); • до 500–700 тыс. руб. (при комплексном внедрении специализированных продуктов с привлечением консультационных услуг и обучении сотрудников). <p>Таким образом, итоговая стоимость зависит от масштаба работ, выбранного программного решения и наличия ресурсов внутри самой организации.</p>	
---	--

Таблица П5

Объяснения моделей в ответ на вопрос «Почему у бизнесменов возникают финансовые проблемы? Приведи экономический закон»

GigaChat 2.0	YandexGPT 5 Pro
<p>Финансовые проблемы у предпринимателей часто связаны с нарушением экономического закона соответствия доходов расходам. Согласно этому закону, финансовое благополучие обеспечивается тогда, когда доходы стабильно превышают расходы, либо расходы оптимизированы таким образом, чтобы соответствовать доходам бизнеса.</p> <p>Основные причины финансовых проблем</p> <p>Несоответствие доходов и расходов:</p> <ul style="list-style-type: none"> • недостаточная прибыльность продукции или услуг; • высокий уровень затрат на производство или операционные издержки; • низкая рентабельность капитала; • отсутствие финансовой дисциплины. <p>Неправильная оценка рисков</p> <p>Предприниматели иногда недооценивают риски, такие как изменение рыночной конъюнктуры, ухудшение экономической ситуации или появление новых конкурентов. Это ведет к финансовым потерям и проблемам с ликвидностью.</p> <p>Проблемы управления капиталом</p> <p>Недостаточное управление оборотным капиталом может привести к нехватке денежных средств для покрытия текущих обязательств. Важная роль принадлежит эффективному управлению запасами, дебиторской задолженностью и кредиторской задолженностью.</p> <p>Чрезмерная долговая нагрузка</p> <p>Высокие проценты по кредитам или неоптимальное использование заемных средств могут значительно ухудшить финансовую устойчивость предприятия.</p> <p>Закон соответствия доходов расходам: предпринимательская деятельность успешна лишь тогда, когда достигается баланс между притоком денег и обязательствами перед кредиторами, поставщиками и сотрудниками. Нарушение этого баланса неизбежно приведет к возникновению финансовых трудностей</p>	<p>Финансовые проблемы у бизнесменов могут возникать по множеству причин, связанных как с внутренними, так и внешними факторами. Один из ключевых экономических законов, который объясняет многие из этих проблем, – это закон соответствия доходов расходам.</p> <p>Основные причины возникновения финансовых проблем:</p> <ol style="list-style-type: none"> 1. Несоответствие доходов и расходов: <ul style="list-style-type: none"> • когда расходы предприятия превышают доходы, это неизбежно ведет к убыткам и финансовым затруднениям; • неправильное планирование бюджета и отсутствие контроля над расходами способствуют возникновению дефицита средств. 2. Недостаточная прибыльность продукции или услуг: <ul style="list-style-type: none"> • если продукция или услуги не приносят достаточной прибыли, это снижает общую доходность бизнеса и может привести к убыткам; • низкая маржинальность товаров или услуг не позволяет покрывать все затраты и получать прибыль. 3. Высокий уровень затрат на производство или операционные издержки: <ul style="list-style-type: none"> • неэффективное использование ресурсов, высокие затраты на сырье, энергию, аренду и зарплату могут значительно увеличивать расходы предприятия; • отсутствие оптимизации процессов ведет к росту издержек. 4. Низкая рентабельность капитала: <ul style="list-style-type: none"> • неэффективное использование вложенного капитала снижает рентабельность бизнеса и уменьшает его способность генерировать прибыль. 5. Отсутствие финансовой дисциплины: <ul style="list-style-type: none"> • нерегулярный учет финансов, отсутствие контроля над денежными потоками и несоблюдение финансовой дисциплины приводят к хаотичному расходованию средств и финансовым проблемам.

	<p>б. Неправильная оценка рисков: Недооценка возможных рисков, таких как изменение рыночной конъюнктуры, действия конкурентов или экономические кризисы, может привести к непредвиденным убыткам.</p> <p>Экономический закон: Закон соответствия доходов расходам гласит, что для обеспечения финансовой устойчивости и прибыльности предприятия необходимо, чтобы доходы стабильно превышали расходы. Это означает, что:</p> <ul style="list-style-type: none">• бизнес должен эффективно управлять своими финансовыми потоками;• необходимо тщательно планировать бюджет и контролировать все расходы;• важно своевременно оценивать и минимизировать риски, которые могут повлиять на финансовую стабильность. <p>Нарушение этого закона неизбежно приводит к финансовым трудностям, включая нехватку средств для покрытия обязательств, снижение ликвидности и, в конечном итоге, возможную неплатежеспособность предприятия</p>
--	---

ЛИТЕРАТУРА

1. Кулинич А.А. Применение больших языковых моделей в системах поддержки принятия решений. Ч. 1. Модели объяснения и большие языковые модели // Проблемы управления. – 2025. – № 1. – С. 47–65. [Kulinich, A.A. Application of Large Language Models in Decision Support Systems. Part I. Explanation Models and Large Language Models // Control Sciences. – 2026. – No. 1. – P. 41–56.]
2. Инженерная психология и эргономика: учебник для вузов / под редакцией Е.А. Климова, О.Г. Носковой, Г.Н. Солнцевой. – 2-е изд., перераб. и доп. – М: Издательство Юрайт, 2025. – 245 с. [Inzhenernaya psihologiya i ergonomika: uchebnik dlya vuzov/ pod redakciej E.A. Klimova, O.G. Noskovej, G. N. Solncevoj. – 2-e izd., pererab. i dop. – M: Izdatel'stvo Yurajt, 2025. – 245 s. (In Russian)].
3. Hoffman, R., Mueller, S., Klein, G., Litman, J. Metrics for Explainable AI: Challenges and Prospects // arXiv:1812.04608. – 2018. – DOI: <https://doi.org/10.48550/arXiv.1812.04608>
4. 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. – No. 1. – P. 347–349. – URL: <https://aclanthology.org/W18-5444/>
5. 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.]
6. Serrat, O. The Five Whys Technique / In: Knowledge Solutions. Springer, Singapore. – 2017. – P. 307–310.
7. Bosselut, A., Le Bras, R., Choi, Y. Dynamic Neuro-Symbolic Knowledge Graph Construction for Zero-Shot Commonsense Question Answering // Proceedings of the AACL Conference on Artificial Intelligence, AACL-21. – Vancouver, Canada, 2021. – Vol. 35, no. 6. – P. 4923–4931.
8. 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>
9. Likert, R. A Technique for the Measurement of Attitudes // Arch. Psychol. – 1932. – Vol. 7, no. 140.
10. Толстова Ю.Н. Измерение в социологии. – М.: Книжный дом «Университет». – 2009. – 288 с. [Tolstova, Yu.N. Izmerenie v sociologii. – M.: Knizhnyj dom «Universitet». – 2009. – 288 s. (In Russian)]
11. Горошниковая Т.А. Вычислительная социология: сдвиг парадигмы или «эконометрика» социологии // Гуманитарные науки. Вестник Финансового университета. – 2021. – Т. 11, № 1. – С. 37–42. [Goroshnikova, T.A. Computational Sociology: A Paradigm Shift or “Econometrics” of Sociology // Humanities and Social Sciences. Bulletin of the Financial University. – 2021. – Vol. 1, no. 1. – P. 37–42. (In Russian)].
12. Митина О.В. Использование модели логистического уравнения для изучения динамических процессов в психологии и науках о жизни // Моделирование и анализ данных. – 2013. – № 1. – С. 61–78. [Mitina, O.V. Ispol'zovanie modeli logisticheskogo uravneniya dlya izucheniya dinamicheskikh processov v psihologii i naukah o zhizni // Modelling and Data Analysis. – M.: Iz-vo GBOU VO «Moskovskij gosudarstvennyj psihologo-pedagogicheskij universitet». – 2013. – No. 1. – P. 61–78. (In Russian)]
13. Thurstone, L.L., Chave, E.J. The Measurement of Attitude. – Chicago: The University of Chicago Press, 1929.
14. Cronbach, L.J. Coefficient Alpha and the Internal Structure of Tests // Psychometrika. – 1951. – Vol. 16, no. 3. – DOI: [doi:10.1007/bf02310555](https://doi.org/10.1007/bf02310555).
15. GigaChat-2. – URL: <https://giga.chat/> (дата обращения: 14.02.2026). [Accessed February 14, 2026.]
16. YandexGPT 5.1. – URL: <https://ya.ru/ai/gpt> (дата обращения: 14.02.2026). [Accessed February 14, 2026.]



17. *Kitcher, P.* Explanatory Unification and the Causal Structure of the World / In: Scientific Explanation. – Minneapolis: University of Minnesota Press. – 1989. – P. 410–505.
18. *Van Fraassen, B.* The Scientific Image. – Oxford: Oxford University Press, 1980.
19. *Hempel, C.* Aspects of Scientific Explanation and Other Essays in the Philosophy of Science // New York: Free Press, 1965.

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>

© 2026 г. Кулинич А. А.



Эта статья доступна по [лицензии Creative Commons «Attribution» \(«Атрибуция»\) 4.0 Всемирная](https://creativecommons.org/licenses/by/4.0/).

APPLICATION OF LARGE LANGUAGE MODELS IN DECISION SUPPORT SYSTEMS. PART II: Measurement and Assessment of Decision-Maker's Satisfaction

A. A. Kulinich

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

✉ alexkul@rambler.ru

Abstract. This paper is a continuation of the multi-part study devoted to the application of large language models (LLMs) in decision support systems (DSSs). Part II considers the issues of measuring the quality of a hybrid DSS that includes a human decision-maker (DM) and an LLM as an assistant. Many criteria for assessing the quality of a hybrid DSS and an expert assessment procedure are proposed. A key feature of the expert assessment of the system's quality is a structural model of a reasonable expert; this model is used to carry out a quasi-experiment for assessing the hybrid system. The assessment is performed within Five Whys, a well-known problem-solving methodology. An example of the operation and assessment of a hybrid DSS is provided, and the experiment results are analyzed. The new method for measuring quality and assessing the DM's satisfaction, based on the structural model of a reasonable expert and a quasi-experiment, can be useful at the development stage of a DSS that incorporates an LLM. This method yields high-quality assessments of hybrid DSSs while reducing the time and cost of assessment without engaging expert groups.

Keywords: decision support, large language model (LLM), hybrid system, assessment criteria, structural model of a reasonable expert.