

УДК 004.89

ГИБРИДНАЯ МОДЕЛЬ ОПТИМИЗАЦИИ БИЗНЕС-РЕШЕНИЙ НА ОСНОВЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ



В.И. Борисенко

*Ассистент кафедры информационных
технологий СПбГУПТД
borisenko.vi.661@suitd.ru*

В.И. Борисенко

Окончил Санкт-Петербургский государственный университет промышленных технологий и дизайна. Область научных интересов связана с разработкой методов и алгоритмов интеллектуального анализа данных.

Аннотация. Представлена концепция верифицируемой системы поддержки принятия решений на основе больших языковых моделей. Предложена гибридная архитектура, интегрирующая методы RAG и ReAct, модули предобработки, генерации, многоуровневой верификации и адаптации. Формализовано преобразование проектных переменных в показатели эффективности с использованием аналитических моделей, машинного обучения и имитационного моделирования. Подход обеспечивает достоверность, интерпретируемость выводов и применимость в корпоративной среде.

Ключевые слова: большие языковые модели, система поддержки принятия решений, RAG, ReAct, гибридная модель, неструктурированные данные

Введение. Современные организации функционируют в условиях высокой неопределённости внешней среды и непрерывно растущих объёмов информации, значительная часть которой представлена в неструктурированном виде. Тексты обращений клиентов, протоколы совещаний, сообщения в социальных медиа, журналы событий информационных систем содержат ценные сведения, необходимые для обоснованного принятия управленческих решений. Однако эффективное использование этих данных ограничено их разнородностью, субъективностью и сложностью формализации в рамках традиционных аналитических моделей.

Критическим фактором, сдерживающим применение методов машинного обучения в бизнес-аналитике, выступает дефицит репрезентативных обучающих выборок. Сбор и разметка достаточного объёма реальных неструктурированных данных сопряжены с рядом объективных трудностей. Конфиденциальность коммерческой информации не позволяет передавать данные за пределы организации. Редкие, но значимые события (отказы оборудования, мошеннические действия, кризисные ситуации) представлены в исторических записях недостаточно. Ручная аннотация текстов требует привлечения экспертов и характеризуется высокой трудоёмкостью. Указанные ограничения препятствуют построению точных прогнозных моделей и проведению всестороннего сценарного анализа.

Научная новизна настоящего исследования заключается в разработке метода оптимизации бизнес-решений, основанного на интеграции верифицированных синтетических неструктурированных данных, полученных от большой языковой модели (large language model, LLM), в контур поддержки принятия решений. Предлагаемый подход,

развивающий архитектуру [1], дополняет генеративные способности языковых моделей многоуровневой системой контроля, обеспечивающей семантическую согласованность, статистическую правдоподобность и процессную выполнимость создаваемых данных. Это позволяет расширить обучающие выборки и сценарные пространства без потери достоверности, повышая точность прогнозных моделей и обоснованность управленческих решений.

Целью работы является формализация, реализация метода верифицированной генерации синтетических неструктурированных данных на основе LLM для последующей оптимизации бизнес-показателей. Для достижения поставленной цели решаются задачи по анализу существующих подходов к синтезу данных, выявлению их ограничений, разработке архитектуры генерации с контролируемыми свойствами, построению модели влияния синтетических данных на интегральный показатель эффективности.

Структура гибридной модели

Для системного совершенствования бизнес-процессов организации необходимо формализовать процесс принятия решений как задачу оптимизации, в которой входные проектные переменные преобразуются в показатели эффективности. Предлагаемая гибридная модель объединяет аналитические зависимости, методы машинного обучения, обученные большие языковые модели и имитационное моделирование, что позволяет учитывать, как структурированные, так и неструктурированные данные (тексты документов, сообщения, экспертные заключения) и обеспечивать высокую точность предсказания. В основе модели лежит архитектура верифицируемой системы поддержки принятия решений, описанная в [1], которая обеспечивает многоуровневый контроль достоверности генерируемых гипотез и их соответствие бизнес-онтологии.

Гибридная модель описывает преобразование множества проектных переменных $X = (X_1, X_2, \dots, X_N)$ в вектор показателей эффективности $Y = (Y_1, Y_2, \dots, Y_K)$. Такое преобразование реализуется совокупностью функциональных модулей, каждый из которых отвечает за определённый этап обработки данных и формирования оценок. В основу структуры положена архитектура верифицируемой системы поддержки принятия решений, предложенная в работе [1], которая обеспечивает многоуровневый контроль достоверности результатов и их согласованность с онтологией предметной области.

Модуль аналитических моделей и моделей машинного обучения реализует вычисление показателей, для которых могут быть построены явные математические зависимости или обучены статистические модели на основе структурированных исторических данных. К числу таких методов относятся регрессионные модели, аппарат временных рядов, методы классификации и кластеризации. Выходные значения модуля используются как часть вектора Y либо как промежуточные переменные для других компонентов.

Модуль обработки неструктурированной информации на основе LLM выполняет извлечение знаний из текстовых данных, таких как документы, новостные сообщения, отчёты, переписка, и генерацию гипотез о влиянии проектных переменных на показатели эффективности. В его основе лежит когнитивный агент, построенный с применением подходов генерации, дополненной поиском (retrieval-augmented generation, RAG) и парадигмы с синергией размышления и действия (reasoning and acting, ReAct) [2, 3]. Первый из них обеспечивает привязку генерируемых утверждений к внешним источникам, второй отвечает за итеративное уточнение гипотез через чередование этапов рассуждения и запросов к данным. Результатом работы модуля являются не только предсказанные значения показателей, но и структурированное объяснение, содержащее цепочку проверенных рассуждений, ссылки на использованные источники и количественные оценки достоверности.

Модуль имитационного моделирования на основе цифрового двойника проигрывает различные сценарии изменения проектных переменных, оценивая результирующие

значения показателей Y в динамике. Данный модуль особенно важен при проверке логико-процессной выполнимости решений, поскольку он выявляет нарушения регламентов и выход метрик за допустимые границы.

Результаты, полученные из перечисленных выше модулей, поступают в модуль верификации, где проходят три уровня контроля. На первом уровне проверяется синтаксико-семантическая согласованность с онтологией предметной области где выявляются противоречия с формализованными бизнес-правилами и связями сущностей. Второй уровень посвящён фактологической обоснованности. Каждое ключевое утверждение сопоставляется с доверенными источниками из пула контекстов и внешними базами данных (регуляторные документы, рыночные индикаторы). Третий уровень оценивает логико-процессная выполнимость, реализуемый через запуск имитационной модели для оценки реализуемости сценария в рамках существующих процессов. По итогам проверки формируется интегральная оценка достоверности предложенного решения.

На последнем этапе модуль коррекции и адаптации аккумулирует обратную связь от лица, принимающего решения (ЛПР), и результаты верификации. На основе анализа отклонённых решений производится корректировка онтологии, добавление или уточнение правил, настройка порогов срабатывания проверок, обновление тегов доверия источников в пуле контекстов. Таким образом, модель обладает свойством адаптивности к изменениям внешней среды и специфике организации.

Взаимодействие перечисленных компонентов организовано по принципу сквозного контура, происходящего от ввода проектных переменных до выдачи верифицированного решения с полным объяснением. Каждый этап обработки сопровождается фиксацией промежуточных данных, что позволяет в модуле коррекции идентифицировать источник ошибок и целенаправленно обновлять соответствующие параметры. Предложенная структура обеспечивает не только высокую точность предсказаний за счёт интеграции разнородных методов, но и необходимый уровень интерпретируемости, что является критическим условием для применения в корпоративной практике.

Проектные переменные и их влияние на показатели. В рамках гибридной модели центральное место занимает формализация связи между управляемыми параметрами бизнес-решения и результирующими показателями эффективности. Проектные переменные $X = (X_1, X_2, \dots, X_N)$ представляют собой совокупность факторов, находящихся в зоне ответственности ЛПР, и подлежащих целенаправленному изменению для достижения желаемых значений показателей. Классификация проектных переменных, их размерность и допустимые области значений определяются спецификой решаемой бизнес-задачи и уровнем детализации модели.

Типология проектных переменных

Проектные переменные могут быть разделены на три категории в зависимости от их природы и способа задания.

Непрерывные переменные. К данной категории относятся параметры, принимающие значения из непрерывного интервала. Примерами служат объёмы производства, размер инвестиций, цена единицы продукции, норматив запаса, длительность этапа производственного цикла. Непрерывные переменные обычно поддаются градиентной оптимизации и могут быть выражены в натуральных или стоимостных единицах.

Дискретные переменные. В эту группу входят параметры, принимающие конечное или счётное множество значений. К ним относятся выбор поставщика из утверждённого перечня, конфигурация программного обеспечения, тип логистической схемы, категория клиентского сегмента. Оптимизация по дискретным переменным требует применения комбинаторных методов или перебора.

Качественные (структурные) переменные. Данная категория охватывает параметры, не имеющие естественного числового представления и определяющие структуру или

конфигурацию бизнес-процесса. Примерами являются политика управления запасами, формат взаимодействия с клиентом, организационная структура подразделения. В гибридной модели качественные переменные кодируются с использованием онтологии предметной области и могут обрабатываться как дискретные альтернативы.

Для каждой переменной X_i задаётся область допустимых значений, определяемая технологическими, ресурсными или регламентными ограничениями. Совокупность этих областей формирует пространство проектных решений.

Отображение проектных переменных в показатели эффективности

Зависимость каждого показателя эффективности Y_i от проектных переменных представляется в виде:

$$Y_k = f_k(X, U), k = 1, 2, \dots, K$$

где U – вектор внешних факторов, не контролируемых ЛПР (рыночная конъюнктура, действия конкурентов, изменения регуляторной среды).

Функции f_k в гибридной модели не являются единообразными. Они формируются на основе комбинации различных вычислительных подходов, выбор которых определяется структурой зависимости и доступностью данных.

Аналитические зависимости применяются в случаях, когда связь между переменными и показателями может быть выражена явной формулой, вытекающей из технологических или экономических закономерностей. Например, показатель объёма выручки может быть выражен как произведение цены на объём продаж, где последний в свою очередь зависит от цены через функцию спроса. Аналитические модели обладают преимуществом прозрачности и вычислительной эффективности, однако их применимость ограничена ситуациями, где известна точная функциональная форма.

Модели машинного обучения используются, когда зависимость имеет сложный, нелинейный характер, но имеется достаточный объём структурированных исторических данных для обучения. В рамках гибридной модели применяются регрессионные методы, деревья решений, градиентный бустинг, а также нейросетевые архитектуры. Обученные модели позволяют учитывать взаимодействия между переменными и нелинейные эффекты, которые трудно формализовать аналитически. При этом в модуле верификации осуществляется контроль за тем, чтобы предсказания модели не выходили за границы области применимости обучающей выборки.

Оценки на основе больших языковых моделей применяются для показателей, существенно зависящих от неструктурированной информации. LLM извлекает из текстовых источников семантические признаки, влияющие на прогнозируемые значения. Например, показатель потребительского спроса может корректироваться на основе анализа тональности отзывов и новостных сообщений о конкурентах. Важной особенностью является то, что LLM не заменяет аналитические или статистические модели, а дополняет их, обеспечивая учёт контекстной информации, не представленной в структурированных данных.

Результаты имитационного моделирования используются для показателей, зависящих от динамики процесса, наличия обратных связей и временных лагов. Цифровой двойник бизнес-процесса позволяет при фиксированном наборе проектных переменных XX получить траекторию изменения показателей во времени, что особенно важно для оценки устойчивости решений и выявления эффектов, проявляющихся только в динамике.

Учёт внешних факторов через неструктурированные данные

Вектор внешних факторов U в классических моделях часто либо игнорируется, либо вводится в виде упрощённых сценариев. Предлагаемая архитектура позволяет извлекать информацию о текущем состоянии внешней среды из неструктурированных потоков с

использованием LLM. Извлечённые факторы интегрируются в функции f_k следующим образом:

$$Y_k = f_k(X, LLM_{\text{extract}}(D_{\text{ext}}))$$

где $LLM_{\text{extract}}()$ – процедура извлечения количественных и качественных характеристик внешней среды, D_{ext} – совокупность внешних текстовых источников.

Полученные оценки проходят верификацию через сопоставление с несколькими независимыми источниками, что снижает риск ошибок, связанных с недостоверной информацией.

Пример формализации для задачи оптимизации маркетинговой стратегии

Для иллюстрации предложенного подхода рассмотрим задачу оптимизации маркетинговой стратегии. Пусть проектные переменные включают:

- X_1 – бюджет рекламной кампании (непрерывная, тыс. руб.);
- X_2 – набор каналов коммуникации (дискретная, варианты: «социальные сети», «контекстная реклама», «телевидение», «комбинированный»);
- X_3 – выбор определенной группы (качественная, кодируемая через онтологию целевых аудиторий).

Показатели эффективности:

- Y_1 – прирост выручки (тыс. руб.);
- Y_2 – рентабельность инвестиций;
- Y_3 – охват целевой аудитории (тыс. человек).

Функция f_1 для показателя Y_1 формируется комбинированно:

$$Y_1 = f_1^{\text{ML}}(X_1, X_2) + \Delta Y_1^{\text{LLM}}$$

где f_1^{ML} – модель машинного обучения, обученная на исторических данных о кампаниях; ΔY_1^{LLM} – корректировка, полученная на основе анализа текущих новостных потоков (например, информация о выходе нового конкурента на рынок, изменении потребительских предпочтений), извлекаемая LLM и верифицируемая через сопоставление с несколькими источниками.

Показатель Y_2 вычисляется аналитически:

$$Y_2 = \frac{Y_1 - C(X_1, X_2)}{X_1}$$

где $C(X_1, X_2)$ – функция затрат, учитывающая бюджет и специфику каналов.

Показатель Y_3 определяется имитационной моделью, учитывающей динамику распространения рекламных сообщений и поведенческие реакции аудитории, параметризованной через X_2 и X_3 .

Роль формализации в оптимизационном контуре

Введение единообразного представления проектных переменных и функций отображения создаёт основу для постановки и решения задачи оптимизации бизнес-решений. Поскольку функции f_k в гибридной модели могут быть недифференцируемыми, содержать дискретные компоненты и включать вычислительно затратные имитационные блоки, для поиска оптимальных значений X^* применяются методы, не требующие вычисления градиентов. Примерами являются эволюционные алгоритмы, байесовская оптимизация, методы поиска с ограничениями. Каждое вычисление целевой функции

$Y = F(Y_1, \dots, Y_k)$ при фиксированном X представляет собой последовательный запуск всех компонентов модели с фиксацией промежуточных результатов, что позволяет модулю верификации контролировать корректность на каждом этапе.

Предложенная формализация обеспечивает:

- однозначную интерпретацию проектных переменных как управляемых параметров;
- возможность комбинирования разнородных вычислительных методов в рамках единого оптимизационного цикла;
- учёт неструктурированных данных наравне со структурированными;
- прозрачность влияния каждой переменной на итоговые показатели через фиксируемые промежуточные зависимости.

Интегральный критерий эффективности

Вектор показателей $Y = (Y_1, Y_2, \dots, Y_k)$ отражает различные, часто противоречивые аспекты эффективности бизнес-решения. Для целенаправленного поиска оптимальных проектных переменных требуется сведение многокритериальной оценки к скалярному значению, позволяющему сравнивать альтернативы и осуществлять направленный поиск.

Интегральный критерий эффективности $Y = F(Y_1, Y_2, \dots, Y_k)$ выполняет функцию агрегирования, при этом выбор функции F и способа её параметризации определяет структуру предпочтений ЛПР.

Линейная свёртка является наиболее распространённым способом агрегирования благодаря простоте интерпретации и вычислительной эффективности [4]:

$$Y = \sum_{k=1}^K w_k Y_k$$

где w_k – весовые коэффициенты, отражающие относительную важность показателей.

Недостатком линейной свёртки является предположение о полной компенсируемости. Снижение одного показателя может быть неограниченно скомпенсировано ростом другого, что не всегда соответствует реальным бизнес-приоритетам.

Учёт предпочтений ЛПР и настройка критерия

Весовые коэффициенты и параметры функций агрегирования не могут быть заданы априори в отрыве от конкретной бизнес-ситуации. В предлагаемой архитектуре настройка критерия осуществляется интерактивно с использованием модуля коррекции и адаптации. ЛПР предоставляет возможность ранжировать альтернативные решения, генерируемые системой, либо указывать на неприемлемые компромиссы. На основе этих данных решается обратная задача по подбору параметров F , при которых выбранные ЛПР альтернативы оказываются предпочтительнее отвергнутых. Для этого могут применяться методы обучения предпочтений, в том числе с использованием LLM для извлечения неявных предпочтений из текстовых пояснений ЛПР [5].

Интеграция с оптимизационным контуром

После определения интегрального критерия задача оптимизации проектных переменных принимает вид:

$$X^* = \arg \max Y(X) = \arg \max F(f_1(X, U), \dots, f_k(X, U))$$

Особенностью гибридной модели является то, что вычисление $Y(X)$ требует последовательного выполнения всех компонентов гибридной модели с фиксацией промежуточных данных для верификации. Каждое вычисление сопровождается формированием объяснения, что позволяет ЛПР контролировать обоснованность получаемых оценок.

Верификация интегрального критерия

Интегральный критерий, как и другие компоненты модели, подлежит верификации. В модуле верификации проверяется:

- согласованность выбранной функции F с онтологией предметной области, например, допустимость линейной компенсации между показателями;
- устойчивость решений к вариациям весовых коэффициентов через анализ чувствительности, позволяющий выявить области параметров, в которых оптимальное решение меняется скачкообразно.

Результаты верификации критерия представляются ЛПР в виде интерактивных графиков (парето-фронты, диаграммы чувствительности), что повышает доверие к рекомендациям системы.

Заключение. Предложенная гибридная модель обеспечивает формализованное преобразование проектных переменных в показатели эффективности с интеграцией больших языковых моделей для обработки неструктурированных данных. Многоуровневая верификация, используя синтаксико-семантическая, фактологическая, процессная проверки, гарантирует достоверность результатов. Введение интегрального критерия с настраиваемой структурой предпочтений и механизмом адаптации на основе обратной связи позволяет решать задачу оптимизации бизнес-решений с требуемой интерпретируемостью. Предложенный подход создаёт основу для построения доверенных систем поддержки принятия решений в корпоративной среде.

Список литературы

- [1] Борисенко, В. И. Генерации синтетических наборов данных с использованием больших языковых моделей для решения бизнес-задач / В. И. Борисенко, В. И. Пименов // ПРОМТЕХДИЗАЙН. Естественные и технические науки : Сборник статей всероссийской научной конференции молодых ученых с международным участием. Часть 3 (СПбГУПТД, 1-27 апреля 2025 г.). – 2025. – С. 40-42.
- [2] Retrieval-Augmented Generation for Large Language Models: A Survey [Электронный ресурс] / Y. Gao [et al.]. – 2024. – Режим доступа: <https://arxiv.org/abs/2312.10997v5>
- [3] ReAct: Synergizing Reasoning and Acting in Language Models [Электронный ресурс] / S. Yao [et al.]. – 2023. – Режим доступа: <https://arxiv.org/abs/2210.03629v3>
- [4] Nasyuha, A. H. Comparison of WSM and Weight Product Methods With WSM-Score and Vector Approaches / A. H. Nasyuha [et al.] // Sinkron : Jurnal Dan Penelitian Teknik Informatika. – 2025. – №2(9). – PP. 948-956.
- [5] Towards a unified view of preference learning for large language models: A survey [Электронный ресурс] / B. Gao [et al.]. – 2024. – Режим доступа: <https://arxiv.org/abs/2409.02795v5>

Авторский вклад

Борисенко Владислав Игоревич – постановка задачи исследования и определение направления совершенствования бизнес-процессов на основе обработки неструктурированных данных, анализ существующих подходов к обеспечению достоверности и интерпретируемости больших языковых моделей, выявление системных ограничений и пробелов в существующих решениях, разработка архитектуры гибридной модели, описание структуры интегрального критерия эффективности и методов агрегирования показателей.

HYBRID MODEL FOR OPTIMIZING BUSINESS DECISIONS BASED ON LARGE LANGUAGE MODELS

V. I. Borisenko

*Assistant of the Department of Information Technology,
SPbSUITD*

Abstract. This article presents a concept for a verifiable decision support system based on large language models. A hybrid architecture is proposed that integrates RAG and ReAct methods, along with modules for preprocessing, generation, multi-level verification, and adaptation. The transformation of design variables into performance metrics is formalized using analytical models, machine learning, and imitation modeling. The approach ensures the reliability and interpretability of conclusions, as well as applicability in a business context.

Keywords: large language models, decision support system, RAG, ReAct, hybrid model, unstructured data.