

Методология и технологии интеллектуальной обработки многомодальной информации на основе регуляризирующего байесовского подхода в условиях неопределенности

С. В. Прокопчина

*Финансовый университет при
Правительстве РФ*

svprokopchina@mail.ru

Д. Л. Тетеревенков

*Финансовый университет при
Правительстве РФ*

teterevdaniil00@gmail.com

А. Ю. Золотаревский

*Финансовый университет при
Правительстве РФ*

artyom@zolotarevskiy.ru

Н. Н. Олтян

*Финансовый университет при
Правительстве РФ*

nikitaoltyan@mail.ru

М. С. Куприянов

*Санкт-Петербургский
государственный
электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)*

Аннотация. В статье рассматривается методология интеллектуальной обработки многомодальной информации на основе регуляризирующего байесовского подхода (РБП) в условиях неопределенности, малых выборок, неполноты данных и концептуальной нестабильности. Центральная идея работы состоит в формировании единого доверенного контура, в котором измерительные данные, экспертные знания, потоковая информация IoT и результаты генеративных моделей приводятся к согласованному представлению, проходят формальную проверку и оцениваются вероятностными методами. Показано, что платформа «Инфоаналитик 3.0» может выступать практической реализацией такой методологии: она объединяет средства интеграции данных, построения цифровых двойников, поддержки принятия решений, детерминированного структурирования неструктурированных текстов и автоматической оценки качества LLM-компонентов. Предлагаемый подход повышает воспроизводимость, интерпретируемость и достоверность интеллектуальных систем, работающих в условиях значительной информационной неопределенности.

Ключевые слова: интеллектуальная обработка информации, многомодальная информация, регуляризирующий байесовский подход, байесовские интеллектуальные технологии, цифровая платформа, LLM, детерминированное структурирование данных, методологическое сопровождение

I. ВВЕДЕНИЕ

Современные прикладные системы все чаще работают не только с большими массивами данных, но и с разнородной информацией: измерениями, экспертными оценками, текстовыми описаниями, потоками от IoT-устройств, результатами моделирования и внешними знаниями предметной области. Поэтому интеллектуальная обработка данных

(ИОД), ориентированная преимущественно на извлечение закономерностей из наблюдений, должна быть расширена до интеллектуальной обработки информации (ИОИ), где данные и знания рассматриваются как взаимодополняющие источники для принятия решений.

Такой переход особенно важен для сложных организационно-технических систем, в которых наблюдения могут быть неполными, нечеткими, противоречивыми или представлены в разных модальностях. В этих условиях недостаточно построить модель по данным: необходимо обеспечить согласование новых наблюдений с уже накопленными знаниями, экспертными ограничениями и целями управления.

Традиционные методы ИОД хорошо работают при наличии репрезентативных выборок и устойчивых распределений, однако в реальных прикладных задачах часто встречаются малые выборки, уникальные эксперименты, дрейф концепции, нестабильные входные потоки и высокая цена ошибки. Поэтому требуется методологическая основа, которая одновременно учитывает неопределенность, сохраняет возможность интерпретации результата и допускает включение экспертных знаний в вычислительный контур.

К современным требованиям ИОИ относятся интерпретируемость, устойчивость, надежность, прозрачность, воспроизводимость и достоверность получаемых решений. Многие методы машинного обучения и генеративного искусственного интеллекта не обеспечивают эти свойства сами по себе, особенно при работе с неструктурированной информацией и открытыми вариантами ответа. В данной статье предлагается рассматривать регуляризирующий байесовский подход (РБП) как методологическую основу для построения доверенного контура ИОИ [1].

Байесовские интеллектуальные технологии (БИТ) на основе РБП позволяют создавать прикладные системы мониторинга, контроля, поддержки принятия решений и управления, способные объединять данные, знания и

Статья подготовлена по результатам исследований, выполненных за счет бюджетных средств по государственному заданию Финансового университета

экспертные оценки. В рамках предлагаемой архитектуры РБП задает вероятностный уровень оценки и выбора решений, платформа «Инфоаналитик 3.0» обеспечивает технологическую реализацию этого уровня, а LLM-компоненты используются как инструменты формализации, генерации гипотез и подготовки рекомендаций, но не как автономный источник истины.

Цель статьи заключается в разработке связанной методологической и технологической схемы, в которой РБП обеспечивает работу с неопределенностью, платформа «Инфоаналитик 3.0» реализует интеграцию разнородных источников, а LLM-компоненты включаются в систему через детерминированный конвейер структурирования данных и контур метрологической оценки.

Научная новизна предлагаемого представления состоит в объединении трех уровней: методологического уровня РБП, платформенного уровня промышленной интеграции и инструментального уровня LLM. Это позволяет рассматривать генеративные модели не как самостоятельную замену аналитике, а как управляемый компонент интеллектуальной обработки информации.

II. МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ РБП И БИТ

Байесовская парадигма естественным образом решает задачу объединения «жестких» измерений с экспертными оценками, поскольку позволяет учитывать априорные знания и уточнять их по мере поступления наблюдений. Однако при нехватке данных классическое правило Байеса может приводить к переобученным или чрезмерно чувствительным распределениям. Регуляризирующий байесовский подход представляет собой синтез классического байесовского вывода и методов регуляризации, направленных на стабилизацию оценок в условиях неопределенности, малых выборок и неоднородных источников информации [1].

- Учет неопределенности: РБП позволяет явно представлять неопределенность данных, модели и экспертных оценок, что особенно важно для сложных систем, где информация может быть неполной, зашумленной, противоречивой или поступающей из разных модальностей.
- Регуляризация: в РБП регуляризация достигается за счет шкалирования на сопряженных шкалах для числовой информации и лингвистических знаний. Это позволяет согласовывать измерения и экспертные суждения, ограничивать сложность модели и получать устойчивые оценки даже при недостатке наблюдений.
- Интерпретируемость: вероятностная форма результата дает возможность объяснять не только выбранное решение, но и степень уверенности в нем, вклад априорных знаний и чувствительность вывода к новым данным.
- Адаптивность: БИТ на основе РБП могут обновлять модель при поступлении новых

данных, обнаруживать изменение условий функционирования и корректировать вывод при дрейфе концепции.

Таким образом, РБП выступает не только вычислительным методом, но и принципом организации единого смыслового представления информации. Данные, экспертные знания, результаты LLM-генерации и измерительные потоки рассматриваются как разнородные свидетельства о состоянии объекта, которые приводятся к согласованной форме и оцениваются с учетом априорных ограничений, контекста и требований предметной области.

Для практического применения такой подход требует технологической среды, в которой можно хранить исходные данные, описывать экспертные шкалы, подключать датчики и внешние системы, фиксировать результаты LLM-преобразований и выполнять последующую проверку качества. Эту роль в работе выполняет платформа «Инфоаналитик 3.0».

III. ПЛАТФОРМА «ИНФОАНАЛИТИК 3.0» НА ОСНОВЕ РБП

Платформа «Инфоаналитик 3.0» создана как комплексное решение для интеллектуального управления сложными системами. Ее роль в предлагаемой методологии состоит в том, чтобы связать методический уровень РБП с прикладными задачами мониторинга, анализа, диагностики и поддержки принятия решений. Платформа объединяет данные и знания, обеспечивает интеграцию с внешними источниками и формирует единый контур обработки информации.

Методология и информационные технологии РБП многократно реализовывались в виде программных комплексов и цифровых платформ. «Инфоаналитик 3.0» является современной реализацией этих технологий, ориентированной на работу с многомодальными источниками, потоковой информацией, цифровыми двойниками и интеллектуальными сервисами, включая генеративные модели.

Платформа построена на принципах современной серверной архитектуры. Используется гибридный подход, сочетающий монолитные и микросервисные компоненты. Такой выбор позволяет сохранить управляемость ядра аналитики и одновременно обеспечить масштабируемость интеграционных сервисов, подключаемых модулей и внешних вычислительных контуров.

Архитектура платформы включает несколько ключевых слоев. Веб-интерфейс обеспечивает взаимодействие пользователей с системой, API-шлюз предоставляет единую точку входа для внешних интеграций, ядро системы реализует бизнес-логику и аналитические алгоритмы, а слой данных хранит как операционную информацию, так и временные ряды. Такое разделение позволяет поддерживать воспроизводимость аналитических процедур и трассируемость полученных результатов.

Технологический стек платформы включает Python и FastAPI для высокопроизводительного API, SQLAlchemy и Pydantic для работы с данными и валидации схем, MySQL для основных данных, InfluxDB для временных рядов, Nginx в качестве обратного прокси и Docker для контейнеризации. IoT-интеграции реализуются через MQTT и REST API, а взаимодействие с внешними

интеллектуальными сервисами может поддерживаться через OpenAPI и Model Context Protocol.

На базе «Инфоаналитик 3.0» можно создавать цифровые двойники предприятий и других сложных систем [2]. Цифровой двойник в данном контексте рассматривается не только как визуальная копия объекта, но и как структурированное представление его состояния, процессов, факторов риска и управленческих связей. Организационная структура предприятия может отображаться в виде дерева подразделений, где цветовая индикация показывает текущее состояние элементов и помогает быстро выделять зоны, требующие анализа или управленческого воздействия.

Архитектура платформы Инфоаналитик 3.0

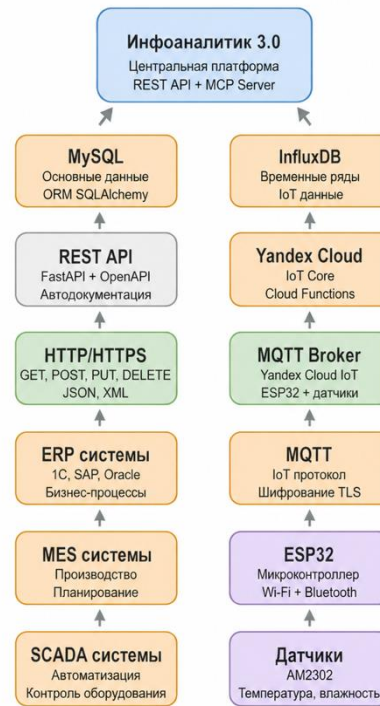


Рис. 1. Архитектура платформы Инфоаналитик 3.0

Схема (рис. 1) демонстрирует многослойную архитектуру платформы. Верхний слой представлен клиентскими приложениями, включая веб-интерфейс и API-клиенты. Центральный слой содержит API-шлюз на базе FastAPI и ядро системы с компонентами бизнес-логики и аналитики. Нижние слои включают хранилища данных, базы временных рядов и инфраструктурные компоненты контейнеризации. Отдельно показана связь с IoT-устройствами и облачными сервисами, что позволяет использовать платформу как единый центр сбора, проверки и интерпретации информации.

Стандартное взаимодействие между клиентом и сервером реализуется через REST API: клиент отправляет HTTP-запросы, сервер обрабатывает их и возвращает структурированные ответы. Такая архитектура упрощает промышленную интеграцию, а формальные схемы данных позволяют фиксировать входы и выходы аналитических процедур, что важно для воспроизводимости и последующей метрологической оценки.

Интеграционные возможности платформы Инфоаналитик 3.0



Поддерживаемые протоколы: Modbus, OPC UA, SNMP, HTTP/HTTPS, MQTT, WebSocket

Рис. 2. Интеграционные возможности платформы «Инфоаналитик 3.0»

Диаграмма (рис. 2) показывает центральную роль «Инфоаналитик 3.0» в интеграции различных систем и протоколов. Платформа связывает REST API-интеграции на базе FastAPI, MQTT-контур для IoT-устройств, MCP-протокол для работы с большими языковыми моделями, а также ERP, MES, SCADA-системы и базы данных. Благодаря этому разнородные источники становятся частью единого информационного пространства.

Реализация API на базе FastAPI обеспечивает автоматическую генерацию документации OpenAPI, валидацию данных на уровне схем, асинхронную обработку запросов и высокую производительность. Для научной и промышленной эксплуатации особенно важно, что такие интерфейсы позволяют фиксировать контракт обмена данными и уменьшать неоднозначность при подключении новых источников.

Ключевые технологии интеграции решают разные задачи: REST API обеспечивает обмен структурированными данными, OpenAPI описывает интерфейсы и схемы, MQTT поддерживает потоковую передачу сообщений от датчиков, а MCP позволяет включать LLM в управляемый контур доступа к инструментам и данным. Тем самым техническая архитектура поддерживает центральную методологическую идею статьи: приведение разнородной информации к единому проверяемому представлению.

Именно на этом уровне возникает методологическая необходимость в строгом управлении LLM-компонентами: они расширяют возможности системы при работе с текстом и знаниями, но требуют формализации, верификации и оценки качества, чтобы результаты генерации могли быть использованы в задачах поддержки решений.

IV. ИНТЕГРАЦИЯ БИТ С LLM

LLM-компоненты могут существенно упростить работу аналитика, снимая рутинные операции первичной формализации текста, подготовки вариантов факторных моделей и генерации пояснительных рекомендаций. В предлагаемой архитектуре LLM не заменяет РБП и экспертную оценку, а выступает как инструмент порождения гипотез и предварительных структур, которые далее проходят проверку, валидацию и вероятностный отбор.

Типовой процесс включает анализ требований, выделение ключевых факторов и подфакторов, построение иерархического дерева, интерактивную доработку модели через диалог с аналитиком и последующую интеграцию данных из внутренних и внешних источников. На каждом этапе результаты LLM должны быть связаны с формальными ограничениями предметной области и с априорными знаниями, представленными в контуре РБП.

Интеллектуальные приборы на уровне периферийных вычислений выполняют первичную обработку и передают агрегированные результаты в облако. Платформа «Инфоаналитик 3.0» выполняет комплексный анализ, машинное обучение и формирование стратегических рекомендаций. Гибридная архитектура объединяет скорость локальной обработки с вычислительными возможностями облака, а интеграции с ERP, MES, SCADA и IoT-устройствами формируют единую экосистему ИОИ.

Одним из ключевых направлений развития платформы является переход от обработки неструктурированной информации к воспроизводимым процедурам ее формализации. Для этого результаты LLM должны быть представлены в виде структурированных объектов, пригодных для проверки схемами, сопоставления с источниками, включения в байесовскую модель и последующего использования в задачах поддержки решений.

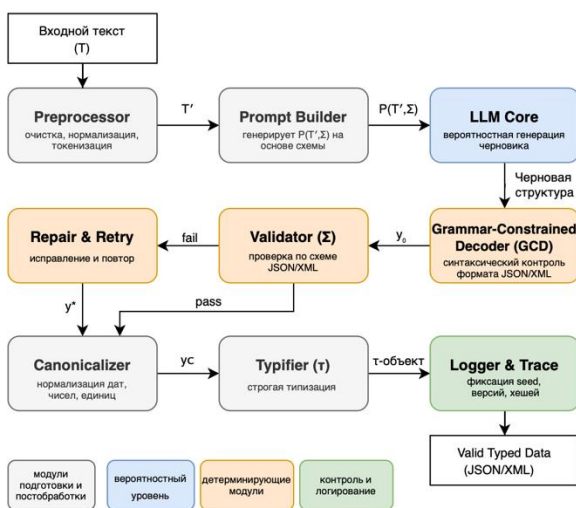


Рис. 3. Архитектура детерминированного конвейера структурирования данных

В рамках предлагаемого подхода генеративные модели рассматриваются не как автономные

элементы, а как компоненты детерминированного конвейера (рис. 3), обеспечивающего управляемое преобразование текстовых данных в структурированные представления. Такой конвейер включает предварительную обработку, формирование запроса, генерацию, синтаксически ограниченное декодирование, валидацию по заданной схеме, исправление отклонений, каноникализацию и приведение типов [3].

Детерминированный уровень обеспечивает синтаксическую корректность и воспроизводимость результата, однако не устраняет содержательную неопределенность, связанную с множественностью допустимых интерпретаций. Поэтому он должен быть дополнен вероятностным уровнем оценки и отбора: LLM формирует альтернативные варианты структурирования, а последующая обработка фильтрует их с учетом априорных знаний, контекста, источников данных и ограничений предметной области.

В результате обработка неструктурированной информации приобретает двухуровневую структуру. На первом уровне обеспечиваются формальная согласованность, валидность схемы и техническая воспроизводимость. На втором уровне выполняется содержательная интерпретация, оценка неопределенности и выбор решения на основе вероятностных критериев РБП. Такое разделение позволяет включать LLM в прикладные системы ИОИ без отказа от требований к достоверности, устойчивости и интерпретируемости.

Проблема оценки LLM заключается в открытом характере генерации: один и тот же запрос может иметь несколько корректных формулировок ответа. Простое сравнение с эталонным текстом не всегда отражает смысловую адекватность, а хорошо сформулированный ответ может оказаться фактически недостоверным, небезопасным или неустойчивым к небольшим изменениям входа. В прикладных сценариях это создает риск скрытых отказов, которые не выявляются одной универсальной метрикой [4].

Поэтому автоматическая оценка качества генерации LLM должна рассматриваться как часть метрологического сопровождения платформы. Ее задача состоит не только в выставлении итогового балла, но и в построении профиля качества, показывающего, какие свойства ответа подтверждены, какие требуют экспертной проверки и где сохраняется содержательная неопределенность.

В состав измеряемых свойств целесообразно включать текстовую близость к эталону, смысловую близость, согласованность с источником, фактическую достоверность, безопасность ответа, устойчивость к возмущениям входа и сценарную полезность. Итоговая оценка не должна сводиться к единственному числу: более корректно формировать вектор качества, отражающий разные стороны поведения системы и пригодный для последующего байесовского анализа.

Связь с РБП здесь принципиальна: показатели качества LLM могут рассматриваться как вероятностные свидетельства о надежности сгенерированного результата. Это позволяет не просто принять или отклонить ответ, а включить его в общий контур вывода с учетом доверия к источнику, полноты данных и требований конкретного сценария.

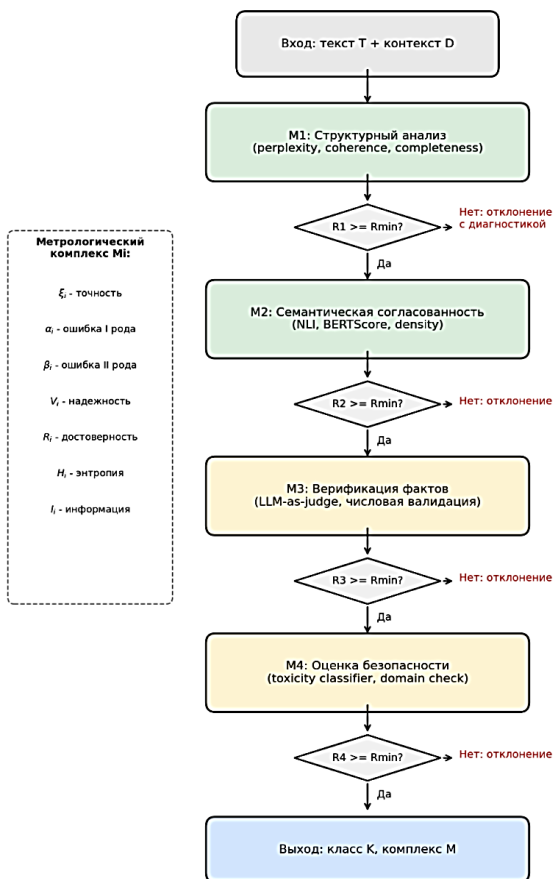


Рис. 4. Контур автоматической оценки качества генерации LLM

Схема (рис. 4) отражает последовательную организацию автоматической метрологии. После получения ответа модели вычисляются лексические и семантические показатели, затем отдельно проверяются фактическая достоверность, безопасность и устойчивость к стресс-тестам. Эти результаты сводятся к уровню сценарной полезности, после чего формируется итоговый профиль качества.

Такой профиль может использоваться в РБП как дополнительное свидетельство при выборе, уточнении или отклонении сгенерированного результата.

V. ЗАКЛЮЧЕНИЕ

В статье предложено единое смысловое представление методологии и технологий интеллектуальной обработки многомодальной информации на основе регуляризирующего байесовского подхода. Показано, что РБП может выполнять роль вероятностного ядра доверенной ИОИ, объединяющего данные, экспертные знания, потоковую информацию, результаты LLM-генерации и формальные ограничения предметной области. Платформа «Инфоаналитик 3.0» рассматривается как практическая реализация этой методологии, обеспечивающая интеграцию разнородных источников, построение цифровых двойников, детерминированное структурирование неструктурированных данных и автоматическую оценку качества генеративных компонентов. Предлагаемый подход может быть рекомендован для создания прикладных интеллектуальных и измерительных систем, функционирующих в условиях информационной неопределенности, многомодальной входной информации, потоковых данных IoT и сенсорных сетей.

СПИСОК ЛИТЕРАТУРЫ

- [1] Прокопчина С.В. Интеллектуальные измерения на основе регуляризирующего байесовского подхода // М.: Научная библиотека, 2021. 497 с.
- [2] Золотаревский А.Ю. Система аналитики и поддержки принятия решений малого производственного предприятия на базе регуляризирующего байесовского подхода // Мягкие измерения и вычисления. 2024. Т. 75, № 2-2. С. 38-48. DOI 10.36871/2618-9976.2024.02-2.004. – EDN SIFQO.
- [3] Олтян Н.Н. Архитектура и алгоритм детерминированного конвейера структурирования JSON и XML данных с использованием LLM // Нейрокомпьютеры: разработка, применение. 2026. № 2.
- [4] Тетеревенков Д.Л. Экспертно-ориентированные методы оценки качества текстовой генерации больших языковых моделей // Мягкие измерения и вычисления. 2025. Т. 90. №5 С.30-37