УДК 004.9
DOI 10.26118/2782-4586.2026.22.77.056

Иванов А.М.
Российский университет дружбы народов
Даутова Д.Т.
Российский университет дружбы народов
Власов Дмитрий Анатольевич
Российский университет дружбы народов

Внедрение программ с продвинутыми языковыми моделями (ИИ) в страховых компаниях

Аннотация. Статья посвящена вопросам внедрения продвинутых языковых моделей (LLM) в страховых компаниях и формированию комплексного подхода к их безопасной и эффективной интеграции. Рассматриваются архитектурные решения, обеспечивающие проверяемость и прозрачность работы моделей, методы оценки экономической эффективности (TCO/ROI) и рисков, нормативные требования в рамках EU AI Act и NIST AI RMF, а также важность долгосрочного управления рисками, связанными с внедрением генеративного ИИ. Особое внимание уделено роли человеческого надзора, независимой валидации и разработке внутренних корпоративных стандартов, обеспечивающих справедливость, надежность и устойчивость решений на базе ИИ. Предлагаемые концепции FRIA-in-the-Loop и LLM-Assurance for Insurance служат основой для формирования целостной системы гарантий качества при использовании генеративных технологий в страховой отрасли.

Ключевые слова: искусственный интеллект, языковые модели (LLM), генеративный ИИ, страхование, автоматизация бизнес-процессов, управление рисками, анализ данных, комплаенс, оценка влияния на фундаментальные права (FRIA), LLM-Assurance for Insurance (LLM-A²), безопасность и справедливость ИИ.

Ivanov A.M.
Peoples' Friendship University of Russia
Dautova D.T.
Peoples' Friendship University of Russia
Vlasov Dmitry Anatolyevich
Plekhanov Russian University of Economics

Implementation of programs with advanced language models (AI) in insurance companies

Abstract. The article focuses on the implementation of advanced language models (LLM) in insurance companies and the development of a comprehensive approach to their safe and effective integration. It explores architectural solutions that ensure the verifiability and transparency of model operations, methods for evaluating cost-effectiveness (TCO/ROI) and risks, regulatory requirements under the EU AI Act and NIST AI RMF, and the importance of long-term risk management associated with the implementation of generative AI. The article emphasizes the role of human oversight, independent validation, and the development of internal corporate standards that ensure the fairness, reliability, and sustainability of AI-based solutions. The proposed FRIA-in-the-Loop and LLM-Assurance for Insurance concepts serve as a foundation for creating a comprehensive quality assurance system for the use of generative technologies in the insurance industry.

Keywords: artificial intelligence, language models (LLM), generative AI, insurance, business process automation, risk management, data analysis, compliance, fundamental rights impact assessment (FRIA), LLM-Assurance for Insurance (LLM-A²), and AI safety and fairness.

Введение

В 2024–2025 годах языковые модели прошли путь от экспериментальной технологии до повседневного производственного инструмента. В страховании это проявилось в автоматизации чтения полисов и актов, составлении резюме по заявкам и кейс-файлам, в помощи андеррайтерам при анализе документов и в диалоговых каналах общения с клиентами. Генеративные модели дополняют классические алгоритмы, действуя как универсальные интерфейсы к документам и базам знаний, а также как оркестраторы функций — запускают расчёты, извлекают реквизиты, формируют черновики ответов и договоров.

Одновременно усилилось внимание регуляторов и общества. В Европейском союзе принят риск-ориентированный подход к системам ИИ: в Регламенте (ЕС) 2024/1689 (EU AI Act) к высокорисковым сценариям отнесены системы для оценки риска и ценообразования по договорам страхования жизни и здоровья; для их пользователей введена обязанность проведения оценки влияния на фундаментальные права (FRIA). В США Национальная ассоциация комиссаров по страхованию (NAIC) приняла модельный бюллетень, задающий ожидания к управлению ИИ, предупреждению недобросовестной дискриминации, мониторингу и доказуемости. Эти документы дополняют уже действующие требования защиты данных (например, право не быть объектом решений, основанных исключительно на автоматизированной обработке, если они влекут юридические последствия).

Следовательно, перед страховщиками стоит двойная задача. С одной стороны — извлечь ощутимую экономическую ценность из LLM (сократить время операций, снизить ошибки и утечки выплат, повысить удовлетворённость клиентов), с другой — обеспечить безопасность, справедливость и юридическую корректность. Цель статьи — предложить системный каркас, позволяющий решать обе задачи: методически связать экономические расчёты с архитектурой решений и процессами управления рисками и комплаенсом.

Основная часть

Нормативный контекст (ЕС, США) и стандарты управления ИИ

EU AI Act вводит риск-стратификацию ИИ-систем и новые обязанности для провайдеров и пользователей высокорисковых решений. Для страхования жизни и здоровья особо значимым является включение систем ИИ для оценки риска и ценообразования в перечень высокорисковых; это влечёт необходимость документации, управления рисками, надзора человеком, требований к качеству данных и кибербезопасности, а также проведение FRIA до ввода в эксплуатацию. В США на уровне штатов формируется практика применения модельного бюллетеня NAIC (12/2023), который закрепляет принципы управления ИИ, требования к мониторингу, документации, управлению третьими сторонами и недопущению недобросовестной дискриминации. [1]

Стандарты позволяют операционализировать требования. NIST AI RMF 1.0 предлагает цикл Govern–Map–Measure–Manage; профиль NIST AI 600-1 фокусирует внимание на рисках генеративных систем (галлюцинации, утечки, злоупотребления). ISO/IEC 42001 описывает систему менеджмента ИИ (AIMS), интегрируемую в общую систему управления качеством организации, а ISO/IEC 23894 — процессный взгляд на риск-менеджмент ИИ. В страховании эти рамки соединяются с отраслевыми методиками актуариев, внутреннего аудита и управлением операционными инцидентами.

Технологические паттерны внедрения LLM

Retrieval-Augmented Generation (RAG) — ключ к проверяемости. Привязка ответов к внутренним документам (полисы, регламенты, акты) и указание источников снижает вероятность «галлюцинаций» и упрощает аудит. Функциональные вызовы и ограниченная агентность превращают модель в оркестратор инструментов: LLM вызывает калькулятор премий, извлекает реквизиты из PDF, заполняет форму FNOL, но не получает неограниченных полномочий.

Ограждения входа/выхода (guardrails) фильтруют персональные и медицинские данные, нормируют стиль и формат ответов, выполняют анти-инъекционную защиту и проверку токсичности/юридических нарушений. Наконец, LLMOps объединяет библиотеку промптов, версионирование, A/B-тесты, журналирование (prompt/response), мониторинг дрейфа и систему выпускных гейтов (quality gates).

Кейсы применения и оценка ценности

Ниже суммированы ключевые кейсы, ожидаемая ценность, рисковая зона и минимальные обязательства комплаенса. Значения выгоды зависят от масштаба и зрелости компании; диаграммы построены на синтетических сценариях и служат для иллюстрации методики отбора. [2]

Таблица 1. Ключевые LLM-кейсы в страховании

Кейс	Цепочка ценности	Оценка выгоды*	Рисковая зона	Комплаенс-минимум
Подсказки андеррайтеру (life/health)	Приём/анализ заявок, реф-чек	Средне–высокая	High-risk (ЕС) → FRIA	FRIA (ст. 27), HITL, документация
Триаж и резюмирование убытков	FNOL, маршрутизация	Средняя	Контекстно значимая	Права на оспаривание, журналы
Документооборот (OCR+RAG)	Бэк-офис, шаблоны	Средняя	Низкая/средняя	Логи, контроль качества данных
Контакт-центр (чат/голос)	Продажи/сервис	Средняя	Требования к прозрачности	Скрипты, фиксация отказов
Антифрод-аналитика	SIU, расследования	Средняя	Повышенные требования fairness	Методология fairness, аудит
ИТ-копайлоты	DevOps/операции	Средняя	Технический риск	Безопасная среда, контроль кода

*Оценки выгоды — ориентировочные; подлежат уточнению на данных конкретной компании.

Диаграмма 1. Оценочный годовой экономический эффект по кейсам (млн $) — синтетические данные

Экономическая модель (EUA)

Для отбора инициатив предложим воспроизводимые формулы. Пусть B — суммарная выгода, TCO — полная стоимость владения, а VaR95 — 95-процентная оценка потенциальных убытков из-за ошибок LLM в конкретном сценарии (например, неверный отказ в выплате, некорректное тарифное предложение, утечка персональных данных).

Выгоды за период T определим как:

$B = \Sigma_i (h_i \cdot c_i + q_i \cdot L_i + r_i \cdot M_i)$,

где h_i — часы, сэкономленные автоматизацией i-го шага; c_i — стоимость часа; q_i — прирост точности/скорости, ведущий к снижению потерь L_i; r_i — рост выручки/удержания; M_i — маржинальный вклад.

Полная стоимость владения:

TCO = (C_лиценз + C_инфра + C_интегр + C_гейткип + C_набл + C_компл) · α_масштаб, где учитываются расходы на модели/API, инфраструктуру (включая кэширование/векторные индексы), интеграцию, ограждения (PII-скрининг, анти-инъекции, выходные фильтры), мониторинг/логгинг и комплаенс (FRIA/DPIA, аудит). Коэффициент α_масштаб отражает рост нагрузки.

Тогда ожидаемая полезность автоматизации (EUA):

EUA = B − TCO − λ · VaR95(ошибки LLM),

где λ — «наказание за риск», выбираемое руководством. Проекты с положительной EUA и лучшей точкой на фронтире риск–доходности получают приоритет. [4]

Безопасность и говернанс: FRIA-in-the-Loop и LLM-A²

FRIA-in-the-Loop — предлагаемая интеграция оценки влияния на фундаментальные права в сам цикл инженерии и эксплуатации. Шаги: (1) идентификация затронутых групп и сценариев вреда; (2) анализ последствий и их вероятностей; (3) проектирование мер — человек-в-контуре, каналы оспаривания, объяснимость; (4) симуляции инцидентов и стресс-тесты; (5) план информирования и компенсаций; (6) периодический пересмотр при изменениях данных/модели/контекста. Это снижает регуляторный и репутационный риск и упрощает аудит.

LLM-Assurance for Insurance (LLM-A²) — минимальный независимый контур верификации страховщика. Он включает техническую секцию (устойчивость к инъекциям и галлюцинациям, верифицируемость, воспроизводимость), юридическую секцию (GDPR/право на оспаривание, потребительское право, обязательства по EU AI Act), актуарную секцию (корректность формул/правил, отсутствие недопустимой дискриминации) и процедуры выпуска (паспорт модели, версия промптов, логирование, лимит «агентности»). [6]

Таблица 2. Риск-регистр LLM и соответствующие контроли

Риск	Описание	Контроли	Отсылка
Автоматизированные решения без HITL	Юридически значимые решения без участия человека	HITL, процедуры оспаривания, логирование	GDPR ст. 22; FRIA
Prompt-инъекция	Вредоносные инструкции внутри данных/документов	Входные фильтры, шаблоны, песочницы инструментов	OWASP LLM Top-10
Дискриминация	Несправедливое обращение с группами	Fairness-аудит, исключение запрещённых/прокси-признаков	Принципы NAIC; EIOPA
Утечки PII/PHI	Экспозиция персональных/медицинских данных	PII-скрининг, шифрование, минимизация контекста	NIST/ISO AIMS
Дрейф/галлюцинации	Нестабильность точности и уверенные ошибки	RAG, re-ranking, верификаторы, выпускные гейты	HELM/Truthful QA; NIST 600-1

Модель зрелости LLMOps для страховщика

Таблица 3. Модель зрелости LLMOps

Уровень	Характеристики	Решения
0 — Ad-hoc	Стихийные эксперименты, нет журнала промптов	Песочницы, запрет ПДн
1 — Пилоты	RAG, базовые гейткиперы	Логи, тесты токсичности
2 — Продукт	CI/CD промптов, метрики, FRIA/DPIA	Паспорт модели, SLO
3 — Масштаб	Портфель инициатив, фронтир риск–доходности	A/B-тесты, персональные гейты
4 — По умолчанию	AIMS (ISO/IEC 42001), синтез данных, непрерывная FRIA	Сквозной аудит, авто-гардрейлы

Кастомизация LLM для страховой отрасли: от общих моделей к отраслевым экспертам

Внедрение LLM в страховании сталкивается с фундаментальным вызовом: общие языковые модели, такие как GPT, Claude или Llama, обучены на разнородных публичных данных и не обладают глубокими знаниями в специфических областях страхования (актуарные расчёты, юридические тонкости полисов, медицинская терминология в health insurance). Это повышает риски «галлюцинаций» и некорректных рекомендаций. Следовательно, следующей эволюционной ступенью является переход к отраслеспецифичным (domain-specific) LLM.

Данный переход может реализовываться по нескольким направлениям:

Дообучение (Fine-tuning) на корпоративных данных: Настройка весов базовой модели на внутренних документах компании: исторических претензиях, полисных условиях, судебных решениях по страховым спорам, протоколах андеррайтинга. Это позволяет модели «усвоить» корпоративный стиль и специфические причинно-следственные связи. Ключевой задачей при этом становится обеспечение качества и репрезентативности данных для обучения, а также постоянное обновление модели по мере изменения продуктов и регуляторной среды.

Разработка страховых эмбеддингов и векторных баз знаний: Создание специализированных векторных представлений для терминов и концепций страхования (например, «франшиза», «co-payment», «subrogation», «оговорка»). Это повышает точность RAG-систем при поиске релевантных фрагментов в документах. Семантический поиск начинает работать не по ключевым словам, а по смыслу, что критически важно для сложных запросов, например, «найти все полисы, где исключается покрытие ущерба от действий третьих лиц при отсутствии вины страхователя».

Интеграция мультимодальности: Страховые кейсы редко ограничиваются текстом. Перспективным направлением является создание моделей, способных анализировать изображения (фото повреждений автомобиля, снимки после пожара, медицинские снимки), аудио (записи звонков в контакт-центр, разбор ДТП) и структурированные данные (таблицы с рисковыми характеристиками, временные ряды выплат). Мультимодальная LLM сможет генерировать комплексный отчёт об убытке, связывая описание клиента, фотографии, данные телематики и исторические прецеденты.

Выбор стратегии зависит от ресурсов, требований к точности и приемлемого уровня риска. Гибридный подход, сочетающий RAG для проверяемости и легкое дообучение для адаптации стиля, представляется оптимальным для большинства страховщиков на текущем этапе.

Экономика LLM: от пилотов к портфельной эффективности

Первоначальный анализ экономической эффективности (EUA) фокусируется на отдельных кейсах. Однако по мере роста числа внедрений возникает задача портфельного управления LLM-инициативами. Это требует более сложных экономических моделей, учитывающих синергии и конфликты между проектами.

Эффект масштаба и синергии инфраструктуры: Развертывание единой LLM-платформы (LLMOps) для всех подразделений (андеррайтинг, урегулирование убытков, сервис) снижает удельные затраты на инфраструктуру, лицензии, мониторинг и комплаенс. Однако необходима четкая система распределения затрат (chargeback), чтобы избежать «трагедии общих ресурсов».

Динамика выгод: Экономический эффект от LLM часто носит нелинейный характер. Например, автоматизация обработки простых заявок (80% потока) может высвободить человеческие ресурсы для анализа сложных и дорогих кейсов (20% потока), где их экспертиза принесет сверхпропорциональную ценность. Это требует расчета не прямого сокращения FTE (Full-Time Equivalent), а увеличения «пропускной способности экспертизы».

Учет долгосрочных и неочевидных издержек:

Стоимость исправления ошибок: Ошибка LLM в андеррайтинге может привести не только к прямому убытку (недосбор премии), но и к необходимости массового пересмотра портфеля и репутационному ущербу.

Стоимость «заморозки» legacy-систем: Интеграция LLM часто происходит поверх существующих унаследованных (legacy) систем. Создание надежных коннекторов и поддержка их работы — скрытая, но существенная статья расходов.

Инвестиции в человеческий капитал: Эффективное использование LLM требует переобучения сотрудников. Страховые андеррайтеры и специалисты по урегулированию убытков должны стать «цифровыми надзирателями», умеющими формулировать промпты, интерпретировать выводы модели и выявлять аномалии. Стоимость таких программ переподготовки должна быть заложена в TCO.

Таким образом, продвинутая экономическая модель должна переходить от расчета EUA для отдельного кейса к построению оптимизационной модели портфеля LLM-проектов, где целевой функцией является максимизация совокупной ценности (B - TCO - Risk Cost) при ограничениях на бюджет, ИТ-ресурсы и регуляторные риски. Методы анализа «что-если» (what-if analysis) и симуляционное моделирование позволяют оценить устойчивость портфеля к изменениям внешней среды (новые регуляторные требования, появление более дешевых моделей, кибератаки).

Регуляторный ландшафт: будущие тренды и проактивная адаптация

Текущие регуляторные рамки (EU AI Act, NAIC Model Bulletin) задают базовые требования. Однако динамика развития технологий и общества указывает на несколько трендов, к которым страховщикам стоит готовиться уже сегодня.

От экзогенного к эндогенному регулированию: Регуляторы будут ожидать, что требования стандартов (NIST AI RMF, ISO 42001) станут не внешним принуждением, а встроенной частью корпоративной культуры и процессов разработки (Security & Compliance by Design). Это означает, что команда data science должна включать в себя специалистов по этике ИИ, юристов и экспертов по комплаенсу с самого начала проекта, а не на этапе аудита.

Фокус на интерпретируемости (Explainability) и причинно-следственном выводе (Causal Inference): Требование «права на объяснение» будет ужесточаться. Для высокорисковых

сценариев (например, отказ в страховании жизни) недостаточно будет указать, на какие данные опиралась модель. Потребуется объяснить логическую цепочку принятия решения. Это стимулирует развитие и внедрение методов контрафактуального анализа («что нужно изменить в анкете, чтобы решение было положительным?») и причинного машинного обучения, что является сложной методологической задачей для «черного ящика» LLM.

Глобальная гармонизация и региональные особенности: В то время как ЕС делает акцент на защите фундаментальных прав (FRIA), а США — на предотвращении недобросовестной дискриминации, другие регионы могут выдвигать свои приоритеты. Например, в Азии может усилиться фокус на кибербезопасности данных и суверенитете ИИ. Страховщикам с глобальным присутствием потребуется разработать модульную систему комплаенса, ядро которой удовлетворяет самым строгим требованиям (де-факто EU AI Act), с адаптируемыми модулями для конкретных юрисдикций.

Регулирование агентских систем (AI Agents): По мере того как LLM превращаются из инструментов генерации текста в автономных агентов, способных совершать цепочки действий (самостоятельно запрашивать справки, вести переговоры о скидках с ремонтными службами), возникнет необходимость в новых регуляторных концепциях. Потребуется определение юридического статуса таких агентов, их ответственности и порядка эскалации инцидентов. Страховым компаниям, внедряющим агентские системы, стоит участвовать в формировании этих норм через отраслевые ассоциации.

Заключение

Внедрение LLM в страховании — это перестройка процессов под управлением норм и стандартов. Компании, которые соединят экономическую дисциплину (модель EUA, портфельный отбор инициатив) с инженерной строгостью (RAG, ограждения, LLMOps) и регулятивной зрелостью (FRIA-in-the-Loop, LLM-A², AIMS), способны перейти от локальных пилотов к устойчивой промышленной эксплуатации. Ключевые принципы успеха: выбор узко очерченных и проверяемых сценариев, безопасность и проверяемость «по умолчанию», независимая валидация и понятная клиенту процедура оспаривания, а также непрерывное улучшение на основании метрик качества и обратной связи. Предложенный каркас обеспечивает согласованность решений на всех уровнях — от архитектуры и экономических расчётов до комплаенса и организационной культуры.

Список источников

1. European Insurance and Occupational Pensions Authority (EIOPA). AI Governance Principles (2021); Digitalization and AI Implementation Reports (2024–2025). — Frankfurt am Main: EIOPA.
2. McKinsey & Company. State of AI 2024 Report: The Dynamics of Generative AI Adoption in Organizations. — New York: McKinsey Global Institute, 2024.
3. National Association of Insurance Commissioners (NAIC). Model Bulletin on Artificial Intelligence. — December 2023. — Principles of NAIC (2020) and state implementation materials (2024–2025).
4. National Institute of Standards and Technology (NIST). AI Risk Management Framework 1.0. — Gaithersburg, MD: NIST, 2023. NIST AI 600-1 — Generative AI Profile. — Gaithersburg, MD: NIST, 2024.
5. OWASP Foundation. OWASP Top-10 for LLM Applications. — Last revisions 2023–2024. — URL: https://owasp.org (дата обращения:19.10.2025).
6. Regulation (EU) 2016/679 — General Data Protection Regulation (GDPR). — Article 22; European Data Protection Board (EDPB) Guidelines on automated decision-making and profiling.

7. Regulation (EU) 2024/1689 of the European Parliament and of the Council on Artificial Intelligence (EU AI Act) // Official Journal of the European Union. — 2024. — 12 июля. — Annex III, ст. 27.
8. Stanford University. HELM — Holistic Evaluation of Language Models; TruthfulQA Benchmark. — URL: https://crfm.stanford.edu/helm/ (дата обращения: 19.10.2025).
9. International Organization for Standardization (ISO); International Electrotechnical Commission (IEC). ISO/IEC 42001:2023 Artificial Intelligence Management System (AIMS); ISO/IEC 23894:2023 Risk Management of AI. — Geneva: ISO, 2023.
10. Industry Examples of Practice: AXA Secure GPT (Azure OpenAI); Allianz “Insurance Copilot”; Zurich ChatGPT experiments in claims settlement (2024).

Сведения об авторах

Иванов А.М., студент Высшей школы управления, Российский университет дружбы народов, Москва, Россия
Даутова Д.Т., студентка Учебно-научного института сравнительной образовательной политики, Российский университет дружбы народов, Москва, Россия
Власов Дмитрий Анатольевич, кандидат педагогических наук, доцент; доцент кафедры математических методов в экономике Российского экономического университете им. Г.В. Плеханова

Information about the authors

Ivanov A.M., student Faculty of Mathematical Modeling of Engineering and Economic Systems Peoples' Friendship University of Russia, Moscow, Russia
Dautova D.T., student Educational and Research Institute of Comparative Educational Policy, Peoples' Friendship University of Russia, Moscow, Russia
Vlasov Dmitry Anatolyevich, Candidate of Pedagogical Sciences, Associate Professor; Associate Professor of the Department of Mathematical Methods in Economics at the Plekhanov Russian University of Economics, Moscow, Russia