УДК 004 URL: https://ptsj.bmstu.ru/catalog/icec/insec/1105.html

МЕТОДИКА ОЦЕНКИ РИСКОВ УТЕЧКИ ДАННЫХ ПРИ ИСПОЛЬЗОВАНИИ LLM В КОРПОРАТИВНОЙ СРЕДЕ: ОРГАНИЗАЦИОННО-ТЕХНИЧЕСКИЕ И ПРАВОВЫЕ АСПЕКТЫ

Д.А. Печенов[1] danilpechenov@gmail.com
SPIN-код: 2998-6946

А.Р. Нигматуллин[2] adel.railevich@yandex.ru
SPIN-код: 4485-1950

[1] *МГТУ им. Н.Э. Баумана, Москва, Российская Федерация*

[2] *Саратовская государственная юридическая академия, Саратов, Российская Федерация*

Работа посвящена проблеме оценки рисков утечки конфиденциальной информации при интеграции больших языковых моделей (БЯМ) в корпоративные бизнес-процессы. Авторы исследуют существующие подходы к защите данных в среде использования искусственного интеллекта (ИИ), выявляют критические векторы утечки информации через пользовательские запросы (промпты) и анализируют пробелы в нормативно-правовом регулировании данной сферы. Особое внимание уделено отсутствию системных методик оценки рисков, что приводит к субъективности принятия решений о допустимости обработки конфиденциальных данных в БЯМ. В статье предложена оригинальная матрица рисков, учитывающая тип обрабатываемых данных, уровень доступа пользователей и архитектуру развертывания моделей (облачное или локальное). Проведенное исследование позволило разработать методику классификации пользовательских запросов по уровню риска с набором правил фильтрации промптов, формулируются рекомендации по применению предложенного подхода в организациях, обрабатывающих персональные данные и коммерческую тайну.

Ключевые слова: большие языковые модели, информационная безопасность, утечка данных, оценка рисков, корпоративная безопасность, матрица рисков, конфиденциальная информация

Введение. Актуальность темы исследования обусловлена стремительным внедрением технологий искусственного интеллекта (ИИ), в частности, больших языковых моделей (БЯМ), в корпоративные бизнес-процессы. Современные организации все чаще используют БЯМ для автоматизации рутинных процессов: анализа документов, генерации отчетности, обработки запросов клиентов и отладки программного кода. Согласно исследованиям рынка, объем инвестиций в технологии ИИ в корпоративном секторе ежегодно увеличивается на 25…30 %, что свидетельствует о высокой востребованности данных решений [1].

 ISSN 2541-8009

Однако широкое применение БЯМ в обработке корпоративной информации сопровождается значительными рисками информационной безопасности. Ключевая проблема заключается в том, что пользовательские запросы (промпты), передаваемые в модель, могут содержать конфиденциальные данные: персональную информацию о сотрудниках и клиентах, коммерческую тайну, финансовую отчетность, техническую документацию и другую защищенную информацию. В случае обработки таких данных в ненадежной облачной среде существует угроза их несанкционированного распространения, хранения на сторонних ресурсах или использования этих данных для формирования вектора атаки на сетевую инфраструктуру компании.

На сегодняшний день в области информационной безопасности отсутствуют системные подходы к оценке рисков, связанных с использованием БЯМ в корпоративной среде. Существующие стандарты и методики, такие как методологии оценки рисков на основе ГОСТ Р 50739–95, ГОСТ Р ИСО/МЭК 15408-1–2012, методы анализа угроз STRIDE, OCTAVE, не учитывают специфику работы с языковыми моделями и не предоставляют инструментов для анализа рисков на уровне пользовательских запросов [2–5]. При использовании ИИ-инструментов организации вынуждены полагаться на общие политики, которые часто носят рекомендательный характер и не обеспечивают должного уровня защиты.

Кроме того, нормативно-правовое регулирование в области защиты персональных данных и коммерческой тайны, в частности, Федеральный закон от 27 июля 2006 г. № 152-ФЗ «О персональных данных» и Федеральный закон от 29 июля 2004 г. № 98-ФЗ «О коммерческой тайне», не содержат конкретных требований к использованию технологий ИИ [6, 7]. Это создает правовую неопределенность и затрудняет разработку решений для безопасного внедрения БЯМ. Таким образом, актуальность исследования определяется необходимостью разработки и нормативного закрепления методики для системной оценки рисков утечки конфиденциальной информации при использовании больших языковых моделей в корпоративной среде.

Основные архитектуры развертывания сети. Современные большие языковые модели могут быть реализованы в двух основных архитектурах развертывания — облачной и локальной. Облачные решения представляют собой публичные сервисы, предоставляемые провайдерами искусственного интеллекта через Интернет. Пользователи взаимодействуют с моделью через веб-интерфейс или интерфейс прикладного программирования API, передавая запросы на удаленные серверы провайдера. Ключевыми преимуществами данного подхода являются минимальные затраты на инфраструктуру, обновление моделей и высокая производительность. Однако облачная архитектура сопря-

жена с рядом рисков в вопросах информационной безопасности. Во-первых, пользовательские запросы и ответы модели передаются по сети и обрабатываются на сторонних ресурсах, что создает угрозу перехвата или несанкционированного доступа к данным. Во-вторых, провайдеры могут сохранять историю взаимодействий для целей улучшения сервиса или обучения моделей, что потенциально приводит к утечке конфиденциальной информации. В-третьих, юрисдикция серверов провайдера может не соответствовать требованиям национального законодательства в области защиты персональных данных.

Локальные решения предполагают развертывание больших языковых моделей на инфраструктуре организации. Модель функционирует в изолированной среде, что исключает передачу данных за пределы корпоративной сети. Данный подход обеспечивает полный контроль над обработкой информации, возможность настройки параметров безопасности и соответствие требованиям нормативных актов и политик компании. Однако локальное развертывание БЯМ требует значительных вычислительных ресурсов, а также специализированных компетенций персонала для администрирования и дообучения моделей. Кроме того, даже при локальном развертывании сохраняется риск внутренних угроз: несанкционированного доступа злоумышленников к чувствительным данным при проникновении во внутреннюю сеть компании или уязвимостей в программном обеспечении.

Таким образом, полноценная защита чувствительной информации при использовании больших языковых моделей возможна исключительно при их локальном развертывании. Облачные решения, несмотря на свою технологическую привлекательность и удобство использования, принципиально не могут обеспечить необходимый уровень конфиденциальности при обработке персональных данных, коммерческой тайны или иной защищенной информации. Передача запросов на сторонние серверы всегда сопряжена с риском их сохранения, анализа или использования в целях, не контролируемых организацией-пользователем. Даже при наличии соглашений о неразглашении и политик конфиденциальности со стороны провайдера организация несет полную ответственность за утечку данных в соответствии с действующим законодательством, в то время как технический контроль над обработкой информации отсутствует [8].

Особенности внедрения больших языковых моделей в корпоративную среду. Культура применения БЯМ для решения рутинных корпоративных задач находится на начальной стадии развития как в мировой практике, так и в Российской Федерации. Согласно исследованиям аналитических агентств, около 70 % российских компаний на 2025 год активно используют технологии ИИ в повседневной деятельности, при этом большинство из них

 ISSN 2541-8009

ограничивается пилотными проектами или обработкой неперсонализированной информации [9]. В то же время наблюдается устойчивый тренд на расширение сфер применения БЯМ: от автоматизации документооборота и анализа текстовых данных до поддержки принятия управленческих решений и генерации контента. Такое стремительное развитие БЯМ создает объективную потребность в разработке методологий обеспечения информационной безопасности, адаптированных к специфике работы с языковыми моделями. Однако на сегодняшний день в данной области существует ряд нерешенных проблем, которые необходимо устранить для обеспечения безопасного внедрения БЯМ в корпоративную среду.

Во-первых, отсутствует системная методика оценки рисков утечки конфиденциальной информации. Организации вынуждены полагаться на общие подходы к анализу угроз, не учитывающие специфику обработки данных через пользовательские запросы к языковым моделям. Существующие методологии (такие как ГОСТ Р 57580, STRIDE, OCTAVE) ориентированы на традиционные информационные системы и не предоставляют инструментов для оценки рисков на уровне промптов.

Во-вторых, наблюдается субъективность принятия решений о допустимости использования БЯМ в конкретных бизнес-процессах. Ответственность за определение границ применения моделей возлагается на сотрудников без четких критериев оценки. Это приводит к несогласованности политик безопасности внутри организации и повышает вероятность ошибок при работе с конфиденциальной информацией.

В-третьих, существует проблема интеграции БЯМ с традиционными системами защиты информации. Большинство решений класса DLP, системы контроля доступа и мониторинга сетевого трафика не учитывают особенности работы с БЯМ и не способны корректно идентифицировать рискованные сценарии использования. Это приводит к необходимости ручной настройки правил или полному отсутствию технического контроля над передачей данных в модели.

В-четвертых, отсутствуют четкие регуляторные требования и отраслевые стандарты в области безопасного использования технологий ИИ. Организации вынуждены самостоятельно интерпретировать общие требования законодательства (федеральные законы № ФЗ-152 и № ФЗ-98) применительно к специфике БЯМ, что создает правовую неопределенность и повышает риски привлечения к ответственности в случае инцидентов с утечкой данных.

Методика оценки рисков с учетом особенностей обработки информации в больших языковых моделях. Устранение данных пробелов требует разработки специализированной методики оценки рисков, учитывающей

особенности обработки информации в среде больших языковых моделей, а также создания практических инструментов для ее реализации в корпоративной среде. Для устранения выявленных проблем — отсутствия системной методики оценки рисков, субъективности принятия решений, сложности интеграции с системами защиты и правовой неопределенности — предложена методика оценки рисков утечки конфиденциальной информации при использовании больших языковых моделей с необходимостью ее нормативно-правового закрепления. Методика основана на построении матрицы рисков, учитывающей ключевые факторы, влияющие на уровень угроз при обработке данных через БЯМ. Основными преимуществами предложенного ниже подхода являются системность оценки рисков на основе четко определенных критериев, гибкость применения к различным архитектурам развертывания и сценариям использования, возможность интеграции с существующими системами защиты информации и прозрачность принятия решений о допустимости использования БЯМ в конкретных бизнес-процессах [10].

Итак, матрица рисков строится на основе четырех ключевых факторов, каждый из которых влияет на вероятность и последствия утечки конфиденциальной информации.

Первый фактор — *тип обрабатываемых данных*. Классификация данных по уровню конфиденциальности определяет потенциальный ущерб в случае их компрометации. Предложено выделить четыре уровня: публичные данные — это свободно доступная информация, не требующая защиты; внутренние данные — это информация для служебного использования внутри организации; конфиденциальные данные — это сведения, составляющие коммерческую тайну или персональные данные; секретные данные — это критически важная информация, утечка которой может нанести непоправимый ущерб организации.

Второй фактор — *уровень доступа пользователя*. Квалификация и полномочия пользователя влияют на вероятность ошибочного или злонамеренного раскрытия информации. Предлагается категорировать пользователей БЯМ следующим образом: обычные сотрудники — пользователи без специальной подготовки в области информационной безопасности и информационных технологий (ИТ); специалисты в области ИТ — пользователи, имеющие специальную подготовку в области информационной безопасности и ИТ, и использующих БЯМ для анализа данных или автоматизации процессов; администраторы — сотрудники, ответственные за настройку и администрирование модели.

Третий фактор — *архитектура развертывания модели*. Способ размещения БЯМ определяет уровень контроля организации над обработкой дан-

 ISSN 2541-8009

ных. Облачный сервис предоставляет минимальный контроль, поскольку данные передаются за пределы организации. Локальный сервер обеспечивает средний уровень контроля — данные остаются в корпоративной сети, но возможен несанкционированный доступ.

Четвертый фактор — *контекст использования*. Тип задачи, решаемой с помощью БЯМ, влияет на характер и объем передаваемых данных. Анализ данных предполагает обработку и структурирование существующей информации, что сопряжено с высоким объемом передаваемых данных. Генерация контента создает риск воспроизведения конфиденциальной информации из контекста или обучающего набора. Сжатие текста требует передачи исходного текста в модель, что также несет риски сохранения полного контекста.

На основе перечисленных факторов авторы предлагают следующую матрицу рисков, позволяющую определить уровень угрозы для конкретного сценария использования БЯМ (табл. 1, 2).

Таблица 1. **Шкала баллов для факторов риска**

Фактор	1 балл	2 балла	3 балла	4 балла
Тип данных	Публичные	Внутренние	Конфиденциальные	Секретные
Уровень доступа	Администратор	Специалист ИТ	Обычный сотрудник	Внешний пользователь/контрагент
Архитектура	Локальный сервер	Локальный сервер с ограниченным доступом	Локальный сервер с полным доступом	Облачный сервис
Контекст	Генерация без текста	Сжатие текста	Анализ	Анализ + генерация

Таблица 2. **Итоговая шкала оценки уровня рисков**

Сумма баллов	Уровень риска	Рекомендуемые меры
4–6	Низкий	Разрешено без ограничений
7–-9	Средний	Обязательное логирование запросов и периодический аудит
10–12	Высокий	Предварительная фильтрация запросов, согласование с руководителем
13–16	Критический	Запрещено к использованию либо разрешено только в исключительных случаях с письменного разрешения руководства

Методика классификации промптов. Для автоматизации оценки рисков на уровне пользовательских запросов (промптов) предложена методика классификации промптов, основанная на анализе их содержания. Суть подхода заключается в выявлении индикаторов конфиденциальной информации в тексте запроса до его передачи в модель. Такой подход позволяет заблокировать или модифицировать рискованные запросы до момента обработки, минимизируя тем самым вероятность утечки данных. Классификация запросов осуществляется на основе анализа ключевых слов и паттернов, характерных для различных типов конфиденциальных данных, а также с учетом структуры запроса, включая его длину, форматирование и наличие табличных данных. Дополнительно учитывается контекст предыдущих запросов в рамках одной сессии взаимодействия с моделью, а также метаданные запроса, такие как идентификатор пользователя, временная метка и источник запроса.

Для каждого типа конфиденциальной информации предложено разработать специализированные правила идентификации, основанные на комбинации ключевых слов, регулярных выражений и структурных признаков. В отношении персональных данных к индикаторам относятся полные ФИО в формате «Фамилия Имя Отчество» или их сокращенные варианты, номера паспортов, состоящие из серии и номера (4+6 цифр), идентификационные номера, такие как ИНН (10 или 12 цифр) и СНИЛС (11 цифр с дефисами), контактные данные в различных форматах телефонных номеров (+7, 8, 7) и электронные почтовые адреса, а также адреса проживания с указанием улицы, дома, квартиры и почтового индекса. Для автоматической фильтрации персональных данных могут применяться регулярные выражения, например, для телефонов — шаблон, учитывающий различные форматы записи номеров, для ИНН — шаблон поиска последовательностей из 10 или 12 цифр, а также словари ключевых слов, включающих термины «паспорт», «ИНН», «СНИЛС», «адрес», «телефон» и «почта».

Коммерчески чувствительная информация может идентифицироваться по наличию финансовых показателей, таких как суммы, бюджеты, доходы и расходы, названий контрагентов, партнеров и клиентов в контексте контрактов, условий соглашений, включая цены, сроки и штрафные санкции, внутренних кодов проектов, продуктов и подразделений, а также стратегических планов, бизнес-целей и прогнозов. Фильтрация данного типа информации осуществляется с помощью словарей ключевых слов, включающих термины «контракт», «договор», «бюджет», «доход», «расход», «стоимость», «цена», «штраф» и «пени», а также выявления числовых значений с валютными символами или явным указанием валюты, например, рублей или долла-

 ISSN 2541-8009

ров. Дополнительно анализируются устойчивые словосочетания, такие как «условия сотрудничества», «план продаж» и «стратегия развития».

Техническая информация выявляется по наличию фрагментов исходного кода, включая ключевые слова языков программирования и характерные структуры кода, названий внутренних систем, сервисов и интерфейсов прикладного программирования (API), конфигураций сетевого оборудования с указанием адресов, портов и параметров, а также упоминаний архитектурных решений и схем баз данных. Для идентификации технической информации используются словари ключевых слов, содержащих термины «код», «функция», «класс», «API», «эндпоинт», «конфигурация» и «сервер», анализ синтаксических конструкций, таких как фигурные скобки, отступы и операторы импорта, а также распознавание форматов файлов, включая расширения .py, .java, .js, .conf, .json и .yaml.

Обработка пользовательского запроса перед его передачей в модель осуществляется пошагово. На первом этапе система принимает текстовый запрос от пользователя и выполняет предварительный парсинг, в ходе которого запрос разбивается на токены, а также анализируется его структура, включая длину, форматирование и наличие табличных данных. Далее система последовательно применяет правила фильтрации для каждого типа конфиденциальных данных, выполняя сканирование на наличие персональных данных, проверку на коммерчески чувствительную информацию и выявление технической информации. На основе найденных индикаторов определяется тип данных, после чего с использованием матрицы рисков рассчитывается итоговый уровень риска с учетом уровня доступа пользователя, архитектуры развертывания модели и контекста использования. В зависимости от рассчитанного уровня риска система принимает одно из возможных решений: при низком уровне риска запрос передается в модель без изменений; при среднем уровне риска пользователю выводится предупреждение с описанием потенциальных рисков и требуется явное подтверждение для продолжения обработки; при высоком и критическом уровнях риска запрос отклоняется, а пользователь получает уведомление о недопустимости обработки; в исключительных случаях при критическом уровне риска запрос может быть направлен на ручную модерацию ответственному сотруднику. Все запросы, особенно те, которые имеют уровень риска выше низкого, записываются в журнал аудита с указанием идентификатора пользователя, временной метки, типа обнаруженных данных и принятого системой решения.

Например, пользователь отправляет запрос, содержащий благодарственное письмо для конкретного лица с указанием полного ФИО, контактного

телефона и полного почтового адреса, включая город, улицу, номер дома и квартиры. Система фильтрации идентифицирует в запросе ФИО, соответствующее стандартному формату трех слов, телефонный номер, соответствующий шаблону для российских номеров, а также адресные данные, содержащие ключевые слова «улица», «дом» и «квартира». Поскольку запрос содержит персональные данные, система блокирует его при использовании облачного сервиса или при отправке обычным сотрудником. При локальном развертывании модели администратором система запрашивает подтверждение с выводом предупреждения о потенциальных рисках.

Внедрение предложенной методики оценки рисков в корпоративную среду требует комплексного подхода, который сочетал бы в себе организационные, технические и образовательные меры. На организационном уровне необходимо разработать внутренние регламенты, определяющие допустимые сценарии использования больших языковых моделей для различных категорий данных и пользователей. Регламенты должны базироваться на результатах оценки рисков по предложенной матрице и устанавливать четкие правила принятия решений о разрешении или блокировке запросов.

Техническая реализация методики предполагает интеграцию механизмов фильтрации запросов в инфраструктуру взаимодействия с моделью. При локальном развертывании БЯМ модуль фильтрации может быть внедрен непосредственно в интерфейс пользователя, обеспечивая анализ запросов до их передачи в модель. В случае использования облачных сервисов применяется прокси-шлюз, через который проходят все исходящие запросы, что позволяет осуществлять контроль даже при отсутствии прямого доступа к платформе провайдера. Интеграция с существующими системами защиты информации, в частности, с решениями класса DLP (программный комплекс, предназначенный для защиты конфиденциальных данных компании от утечек), позволяет использовать корпоративные словари конфиденциальных терминов для повышения точности фильтрации и минимизации ложных срабатываний.

Меры по обеспечению функционирования методики. Для обеспечения устойчивости функционирования методики необходимо организовать непрерывный мониторинг ее эффективности. Это включает анализ журналов запросов с целью выявления ложных срабатываний и пропусков рискованных сценариев, а также регулярную актуализацию правил фильтрации в соответствии с изменением бизнес-процессов и появлением новых категорий конфиденциальной информации. Периодический аудит использования БЯМ позволяет своевременно выявлять отклонения от установленных политик и корректировать подходы к управлению рисками.

 ISSN 2541-8009

Предложенная методика обладает достаточной гибкостью для адаптации под особенности организаций различного масштаба и отраслевой принадлежности. Ее реализация не требует радикальной перестройки существующей инфраструктуры информационной безопасности, а может быть осуществлена поэтапно с постепенным наращиванием функциональности механизмов контроля. Такой подход делает методику применимой как в крупных корпорациях с развитыми системами защиты, так и в организациях со средним уровнем зрелости процессов информационной безопасности.

Вместе с тем в целях эффективного применения методики для оценки рисков утечки конфиденциальной информации при использовании БЯМ представляется обоснованным ее нормативное закрепление на различных уровнях правового регулирования. Так, в частности, на уровне федерального законодательства целесообразно установить обязанность применения риск-ориентированного подхода при использовании технологий ИИ, а на уровне подзаконных нормативных актов и методических документов уполномоченных органов (Роскомнадзор, ФСТЭК России и др.) возможно закрепление конкретных процедур оценки рисков, включая использование матриц рисков и механизмов классификации пользовательских запросов. Дополнительно же методика может быть реализована в национальных стандартах и внутренних нормативных актах организаций, что обеспечит ее практическое применение и адаптацию к специфике конкретных бизнес-процессов.

Заключение. Подводя итоги вышесказанному, отметим, что исследование актуальной проблемы обеспечения информационной безопасности при использовании БЯМ в корпоративной среде выявило ряд системных недостатков в существующих подходах к защите конфиденциальной информации. К числу ключевых проблем относятся отсутствие специализированных методик оценки рисков, ориентированных на специфику работы с языковыми моделями, преобладание субъективных решений при определении допустимости использования БЯМ для обработки чувствительных данных, недостаточная адаптация традиционных систем защиты информации к особенностям взаимодействия с ИИ, а также отсутствие четких нормативных требований в данной области. Эти факторы в совокупности создают серьезные барьеры для широкого и безопасного внедрения технологий ИИ в бизнес-процессы организаций.

В качестве решения указанных проблем авторами была разработана методика оценки рисков утечки конфиденциальной информации при использовании БЯМ. Основу методики составляет матрица рисков, учитывающая четыре ключевых фактора: тип обрабатываемых данных, уровень доступа пользователя, архитектуру развертывания модели и контекст использования.

Комбинация данных факторов позволяет количественно оценить уровень риска конкретного сценария использования БЯМ и принять обоснованное решение о допустимости обработки данных.

Дополнительно предложена методика классификации пользовательских запросов, основанная на автоматическом выявлении индикаторов конфиденциальной информации в тексте промптов. Методика включает правила идентификации персональных данных, коммерческой тайны и технической информации с использованием комбинации ключевых слов, регулярных выражений и анализа структурных признаков запроса. Алгоритм обработки запроса предусматривает пошаговую проверку на наличие конфиденциальных данных, оценку риска по предложенной матрице и принятие решения — разрешить, запросить подтверждение, заблокировать или направить на ручную модерацию.

Научная новизна исследования заключается в разработке первой в отечественной практике системной методики оценки рисков, специально адаптированной для работы с БЯМ. В отличие от существующих подходов, ориентированных на традиционные информационные системы, предложенная матрица учитывает специфику обработки данных через пользовательские запросы и позволяет оценивать риски на уровне отдельного промпта. Методика классификации запросов дополняет матрицу практическим инструментом автоматизации оценки, что снижает зависимость от субъективного фактора и повышает воспроизводимость результатов.

Таким образом, предложенная методика представляет собой законченный инструментарий для управления рисками утечки конфиденциальной информации при использовании БЯМ и может стать основой для формирования стандартов безопасного внедрения технологий ИИ в корпоративную среду.

Литература

[1] *Интеллект оправдывает средства: топ-5 IT-компаний России обеспечили 95% выручки на рынке ИИ.* URL: https://www.kommersant.ru/doc/7989450 (дата обращения 22.03.2026).

[2] ГОСТ Р 50739–95. *Средства вычислительной техники. Защита от несанкционированного доступа к информации. Общие технические показатели*. Москва, Изд-во стандартов, 1995.

[3] ГОСТ Р ИСО/МЭК 15408-1–2012. *Информационная технология. Методы и средства обеспечения безопасности. Критерии оценки безопасности информационных технологий.* Москва, Стандартинформ, 2014, ч. IV, 54 с.

 ISSN 2541-8009

[4] *Средство моделирования угроз Майкрософт (Microsoft Threat Modeling Tool)*. URL: https://learn.microsoft.com/ru-ru/azure/security/develop/threat-modeling-tool-threats (дата обращения 22.03.2026).

[5] Alberts C., Dorofee S. *OCTAVE Criteria, Version 2.0. Technical Report CMU/SEI-2001-TR-016.* Pittsburgh, PA, Software Engineering Institute, Carnegie Mellon University, 2001, 109 p.

[6] *О коммерческой тайне*. ФЗ от 29.07.2004 № 98-ФЗ. 2004, № 31, ст. 3217.

[7] *О персональных данных*. ФЗ от 27.07.2006 № 152-ФЗ. 2006, № 31 (ч. 1), ст. 3451.

[8] Самолкаева А.М., Шведова С.М. Искусственный интеллект в сфере информационной безопасности: преимущества, ограничения и перспективы. *Вестник науки*, 2024, № 3(72), с. 534–539.

[9] Генерати*вный искусственный интеллект используют 70 % российских компаний*. URL: https://www.forbes.ru/tekhnologii/535047-generativnyj-iskusstvennyj-intellekt-ispol-zuut-70-rossijskih-kompanij (дата обращения 22.03.2026).

[10] Артамонов В.А., Артамонова Е.В. Искусственный интеллект и безопасность: проблемы, заблуждения, реальность и будущее. *Россия: тенденции и перспективы развития. XXI Нац. науч. конф. с междунар. уч.: матер*. Москва, ИНИОН РАН, 2022, № 17-1.

Поступила в редакцию 27.03.2026

Печенов Данила Александрович — студент кафедры «Безопасность в цифровом мире» МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Нигматуллин Адель Раилевич — аспирант кафедры прокурорского надзора и криминологии Саратовской государственной юридической академии, Саратов, Российская Федерация.

Ссылку на эту статью просим оформлять следующим образом:
Печёнов Д.А., Нигматуллин А.Р. Методика оценки рисков утечки данных при использовании LLM в корпоративной среде: организационно-технические и правовые аспекты. *Политехнический молодежный журнал*, 2026, № 03 (104). URL: https://ptsj.bmstu.ru/catalog/icec/insec/1105.html

A METHOD FOR ASSESSING DATA LEAK RISKS WHEN USING LLM IN A CORPORATE ENVIRONMENT: ORGANIZATIONAL, TECHNICAL, AND LEGAL ASPECTS

D.A. Pechenov[1] danilpechenov@gmail.com
SPIN-code: 2998-6946

A.R. Nigmatullin[2] adel.railevich@yandex.ru
SPIN-code 4485-1950

[1] *Bauman Moscow State Technical University, Moscow, Russian Federation*

[2] *Saratov State Law Academy, Saratov, Russian Federation*

The paper is devoted to the problem of assessing the risks of confidential information leakage when integrating large language models (LLM) into corporate business processes. The authors explore existing approaches to data protection in the environment of artificial intelligence (AI), identify critical vectors of information leakage through user requests (prompta) and analyze gaps in the regulatory framework in this area. Particular attention is paid to the lack of systematic risk assessment methods, which leads to the subjectivity of decision-making on the permissibility of processing confidential data in the LLMs. The article offers an original risk matrix that takes into account the type of data being processed, the level of user access, and the architecture of model development (cloud or local). The conducted research made it possible to develop a methodology for classifying user requests by risk level with a set of rules for filtering prompt, and recommendations are formulated for the application of the proposed approach in organizations that process personal data and trade secrets.

Keywords: large language models, information security, data leakage, risk assessment, corporate security, risk matrix, confidential information

Received 27.03.2026

Pechenov D.A. — Student of the Department of Security in Digital World of Bauman Moscow State Technical University, Moscow, Russian Federation.

Nigmatullin A.R. — Postgraduate student at the Department of Prosecutorial Supervision and Criminology of Saratov State Law Academy, Saratov, Russian Federation.

Please cite this article in English as:
Pechenov D.A., Nigmatullin A.R. A method for assessing data leak risks when using LLM in a corporate environment: organizational, technical, and legal aspects. *Politekhnicheskiy molodezhnyy zhurnal*, 2026, no. 03 (104). (In Russ.). URL: https://ptsj.bmstu.ru/eng/catalog/icec/insec/1105.html

 ISSN 2541-8009