

Результаты опроса Ассоциации «Институт внутренних аудиторов»: риски использования общедоступного ИИ

Период опроса: 11-17 марта 2025 года

Участники опроса: 211 внутренних аудиторов российских компаний и организаций

Вопрос 1. Выберите наиболее значимые, по вашему мнению, риски использования общедоступного ИИ (например: Gigachat, YaGPT, ChatGPT, DeepSeek и прочее) для рабочих целей (можно было выбрать несколько вариантов ответов):

1. Риск утечки конфиденциальных данных/ Нарушение конфиденциальности данных – **73%**
2. Ошибки и недостоверная информация/ Создание дезинформации / Галлюцинирование (модели могут генерировать неполные или ложные ответы, а также выдавать правдоподобную, но фактически ложную информацию) – **70%**
3. Непонимание принципов работы ИИ сотрудниками (как результат: некорректное использование возможностей ИИ или неспособность выявить ошибки) – **43%**
4. Избыточная зависимость от технологий (сотрудники перестанут критически оценивать данные и решения, выработают привычку полагаться на ИИ, а также утратят креативное мышление) – **40%**
5. Непреднамеренное нарушение законодательства (ответы ИИ могут не соответствовать актуальным законодательным нормам) – **39%**
6. Сложность валидации вывода \ неаудируемость (сложность проверки точности данных или источников данных, которые генерирует и использует модель) – **38%**
7. Человеческий фактор в обработке данных при передаче в ИИ, как результат: некорректные выводы – **30%**
8. Риски злоупотреблений и мошенничества/ Риски манипуляции контентом (злоумышленники могут создавать фальшивые запросы или автоматизированные модели для фишинга, а также для генерации мошеннических писем, фейковых новостей или ложной информации) – **29%**
9. Дискриминация и предвзятость (модели могут отражать предвзятости, заложенные в данных обучения) – **22%**
10. Размытие ответственности / Правовые конфликты с поставщиками технологий (споры вокруг ответственности за ошибки ИИ между пользователем и поставщиком технологии) – **18%**
11. Риск потери рабочих мест (замена сотрудников ИИ-агентами) – **14%**

12. Низкая предсказуемость и воспроизводимость результатов (генерация ответов статистически, а не на основании четкой логики) – 13%
13. Негативное воздействие на репутацию компании (если ИИ выдает некорректную или оскорбительную информацию, которая становится публичной) – 12%
14. Культурные и языковые барьеры (модели, созданные преимущественно на данных других языков, могут недостаточно понимать контекст на языке запроса) – 12%
15. Нарушение авторских прав (при генерации контента) – 11%
16. Этические конфликты (использование ИИ для автоматизации некоторых процессов может восприниматься как неэтичный подход) – 8%
17. Другое:
 - Нужно будет больше рабочих мест, чтобы верифицировать данные ИИ
 - Для корпоративной работы нельзя использовать общедоступный ИИ
 - Некорректный сбор информации и данных для аудита (некорректные риск-ориентированные выборки для исследования)

Вопрос 2. По вашему мнению, какие меры могут снизить риски, сопровождающие использование открытых ИИ-инструментов в рабочих целях? (можно было выбрать несколько вариантов ответов)

1. Обучение сотрудников (о рисках использования публичных языковых моделей, о возможностях моделей, о правилах работы и т.д.) – 60%
2. Создание внутренних политик использования ИИ – 57%
3. Ограничение предоставления конфиденциальной информации (ограничение отправки в публичные языковые модели конфиденциальных и/или чувствительных данных) – 55%
4. Верификация источников (проверка информации, предоставляемой моделью, с внешними надежными источниками данных) – 53%
5. Разделение задач (использование публичных языковых моделей только для общих запросов, а для работы с чувствительными данными – только внутренние решения) – 45%
6. Периодический анализ рисков и внутренний аудит (регулярный пересмотр политики использования ИИ и проведение аудита процедур для оценки новых рисков) – 44%
7. Развертывание моделей только во внутреннем изолированном контуре (избегание использования облачных решений) – 38%
8. Использование промежуточных API или шлюзов безопасности (для всех обращений к модели) с последующим мониторингом использования – 29%
9. Лимит и контроль доступа (доступ к публичным языковым моделям только у тех сотрудников, которым это необходимо для выполнения рабочих задач) – 24%
10. Резервные проверки ответов (процессы проверки текста \ выводов, сгенерированного языковыми моделями) – 24%

11. Фильтрация выходных данных модели (программные фильтры или механизмы пост-обработки для анализа генерируемого текста на соответствие правилам) – 22%
12. Юридическая оценка использования технологий (юридические консультации по использованию модели в контексте конфиденциальности, авторских прав, ответственности за генерируемый контент и локальных правил использования ИИ) – 21%
13. Шифрование данных (запросы и ответы от публичных языковых моделей шифруются для защиты от потенциального перехвата данных) – 17%
14. Использование кастомизированных облачных решений (использование специализированных версий языковых моделей (например, корпоративных решений OpenAI или других провайдеров)) – 12%
15. Другое (свободное заполнение):
 - Тестирование сотрудников после обучения и регулярно в работе в области применения ИИ
 - Разработка инструментов ИИ для холдинга/ компании (уход от продуктов общего пользования)
 - Создание своих внутренних ИИ-помощников на основании внешних ИИ-помощников, адаптация для своего внутреннего контура
 - Если модель самообучающаяся, то ошибок и невозможности расшифровки данных не избежать на всем периоде ее использования
 - Использование верифицированных источников данных
 - Коллективная экспертная верификация ответов

Вопрос 3. Какие меры использует ваша компания для снижения рисков, сопровождающих использование открытых ИИ-инструментов в рабочих целях? (можно было выбрать несколько вариантов ответов)

1. Ограничение предоставления конфиденциальной информации (ограничение отправки в публичные языковые модели конфиденциальных и/или чувствительных данных) – 46%
2. Создание внутренних политик использования ИИ – 23%
3. Обучение сотрудников (о рисках использования публичных языковых моделей, о возможностях моделей, о правилах работы и т.д.) – 22%
4. Развертывание моделей только во внутреннем изолированном контуре (избегание использования облачных решений) – 18%
5. Периодический анализ рисков и внутренний аудит (регулярный пересмотр политики использования ИИ и проведение аудита процедур для оценки новых рисков) – 17%
6. Лимит и контроль доступа (доступ к публичным языковым моделям только у тех сотрудников, которым это необходимо для выполнения рабочих задач) – 16%
7. Верификация источников (проверка информации, предоставляемой моделью, с внешними надежными источниками данных) – 15%
8. Разделение задач (использование публичных языковых моделей только для общих запросов, а для работы с чувствительными данными – только внутренние решения) – 13%

9. Использование промежуточных API или шлюзов безопасности (для всех обращений к модели) с последующим мониторингом использования – **11%**
10. Резервные проверки ответов (процессы проверки текста/ выводов, сгенерированного языковыми моделями) – **9%**
11. Фильтрация выходных данных модели (программные фильтры или механизмы пост-обработки для анализа генерируемого текста на соответствие правилам) – **7%**
12. Юридическая оценка использования технологий (юридические консультации по использованию модели в контексте конфиденциальности, авторских прав, ответственности за генерируемый контент и локальных правил использования ИИ) – **6%**
13. Шифрование данных (запросы и ответы от публичных языковых моделей шифруются для защиты от потенциального перехвата данных) – **6%**
14. Использование кастомизированных облачных решений (использование специализированных версий языковых моделей (например, корпоративных решений OpenAI или других провайдеров)) – 5%
15. Другое (свободное заполнение):
 - Ничего не используется **(15%)**
 - Компания не использует ИИ **(10%)**
 - Внутренний контроль
 - Резервные проверки и верификация проводятся в «ручном режиме», тем самым увеличиваются трудозатраты

Вопрос 4. По вашему мнению, достаточное ли внимание ваша компания уделяет управлению рисками, связанными с использованием ИИ в рабочих целях?

- Достаточно, регулярно используем все основные лучшие практики – **9%**
- В целом достаточно, но есть куда расти/ требуется доработка – **22%**
- Уделяем этим рискам минимальное достаточное внимание – **18%**
- Недостаточно/ не управляем данными рисками в качестве регулярного осознанного процесса – **33%**
- Затрудняюсь ответить – **23%**

Справка

Ассоциация «Институт внутренних аудиторов» (Ассоциация «ИВА»), зарегистрированная в 2000 г., является профессиональным объединением внутренних аудиторов, внутренних контролеров и работников других контрольных подразделений российских компаний и организаций.

www.iaa-ru.ru