rocket lab

команда роста на аутсорсе
для системного улучшения ключевых метрик

# Шаблон политики безопасности при использовании ИИ в компании

Готовый документ для безопасного и этичного внедрения ИИ-инструментов

## Цель и область применения

Настоящий документ определяет правила и принципы безопасного, этичного и эффективного использования инструментов искусственного интеллекта (ИИ) в [Название Вашей Компании].

Цель:

* Защита конфиденциальных данных и коммерческой тайны компании.

* Предотвращение рисков, связанных с кибербезопасностью и этическими аспектами ИИ.

* Повышение эффективности бизнес-процессов за счет ответственного внедрения ИИ-инструментов.

* Формирование культуры ответственного использования ИИ среди сотрудников.

Область применения:

Данная политика обязательна для всех сотрудников [Название Вашей Компании], временных работников, подрядчиков и любых третьих лиц, имеющих доступ к корпоративным системам и данным, использующих ИИ-инструменты в рамках своей профессиональной деятельности. Политика распространяется на все ИИ-инструменты, как внутренние, так и внешние (публичные), используемые для рабочих задач.

## Ключевые принципы использования ИИ

- ☐ Конфиденциальность данных

  Категорически запрещается вводить коммерческую тайну, персональные данные клиентов или сотрудников, финансовые отчеты, фрагменты кода, не подлежащие публичному раскрытию, и любую другую конфиденциальную информацию в публичные ИИ-системы, не авторизованные компанией.

- ☐ Точность и проверка информации

  Сотрудник обязан проверять факты, данные и любой контент, сгенерированный ИИ, перед его использованием или публикацией. ИИ может допускать ошибки (галлюцинации), и ответственность за достоверность конечного результата лежит на пользователе.

- ☐ Этичность и предотвращение предвзятости

  ИИ-инструменты должны использоваться таким образом, чтобы не допускать дискриминации, предвзятости или создания контента, который может быть расценен как оскорбительный, неэтичный или нарушающий внутренние нормы компании и законодательство.

- ☐ Ответственность пользователя

  Каждый сотрудник несет персональную ответственность за соблюдение данной политики и последствия использования ИИ-инструментов в своей работе.

- ☐ Прозрачность использования ИИ

  В случае, если контент или решение были сгенерированы ИИ, это должно быть явно указано, особенно при взаимодействии с клиентами или внешними партнерами.

## Классификация ИИ-инструментов и правила использования

| Категория | Примеры ИИ-инструментов | Условия использования |
|---|---|---|
| Разрешенные к использованию | YandexGPT, GigaChat, Kandinsky, Битрикс24 CoPilot, AmoCRM (встроенные ИИ-функции), Unisender (ИИ для текстов), Yandex DataLens, Kaiten (ИИ-анализ задач), Happy Job (HR-аналитика), Albato (автоматизация), Testograf (аналитика опросов) | Разрешены для использования со всеми типами данных, кроме особо конфиденциальных (см. 'Запрещенные'). Приветствуется использование для повышения эффективности рутинных задач (автоматизация ответов, генерация черновиков, анализ общедоступных данных). |
| Ограниченные к использованию | ChatGPT, Midjourney (публичные версии) | Допускается использование только для работы с нечувствительными данными, не содержащими коммерческую тайну, персональные данные или другую конфиденциальную информацию компании. Например, для генерации идей, создания общедоступного контента, анализа публичной информации. Обязательна проверка всех сгенерированных данных. |
| Запрещенные к использованию | Любые неавторизованные публичные ИИ-системы для бизнес-критичных задач; ИИ-инструменты, не прошедшие внутреннюю проверку безопасности; ИИ-системы, требующие ввода конфиденциальных данных без явного разрешения руководства. | Категорически запрещено использовать для обработки коммерческой тайны, персональных данных, финансовой информации, фрагментов кода или любой другой конфиденциальной информации. Нарушение может привести к серьезным дисциплинарным мерам. |

## Работа с данными и ИИ-системами

- ☐ Правила ввода данных в ИИ-системы

  Перед вводом любых данных в ИИ-инструмент убедитесь, что они соответствуют категории использования (разрешенные, ограниченные, запрещенные). Никогда не вводите конфиденциальную информацию в публичные ИИ, не предназначенные для этого. Используйте внутренние RAG-системы для доступа к корпоративной информации.

- ☐ Политика хранения и удаления данных

  Данные, используемые ИИ-системами, должны храниться в соответствии с внутренними политиками компании по хранению данных. При использовании внешних ИИ-сервисов, ознакомьтесь с их политикой конфиденциальности и убедитесь, что она соответствует требованиям компании. В случае необходимости, данные должны быть удалены из ИИ-систем в соответствии с регламентом.

- ☐ Использование внутренних RAG-систем

  Для безопасного доступа к корпоративной информации (регламенты, инструкции, каталоги товаров) рекомендуется использовать внутренние RAG (Retrieval-Augmented Generation) системы. Это позволяет языковым моделям обращаться к актуальной и релевантной информации без риска утечки во внешние сервисы.

## Кибербезопасность ИИ

- ☐ Защита от Data Poisoning

  Компания внедряет методологию MLSecOps для обеспечения безопасности на всех этапах жизненного цикла ИИ-моделей. Это включает контроль качества входных данных, регулярный мониторинг поведения моделей и защиту от целенаправленного 'отравления' данных, которое может привести к внедрению бэкдоров или неверным решениям.

- ☐ Контроль доступа к ИИ-агентам и API-ключам

Все ИИ-агенты и API-ключи, предоставляющие доступ к корпоративным системам, должны иметь минимально необходимые права доступа. Доступ должен быть строго ограничен и регулярно пересматриваться. Любая компрометация API-ключа или токена ИИ-агента рассматривается как серьезный инцидент безопасности.

- [ ] Регулярные аудиты и обновления ИИ-систем

  Все используемые ИИ-системы, как внутренние, так и внешние, должны проходить регулярные аудиты безопасности. Программное обеспечение и модели должны своевременно обновляться для устранения выявленных уязвимостей. Особое внимание уделяется уязвимостям интеграционного слоя, который связывает ИИ с внутренними системами компании.

- [ ] Предотвращение Shadow AI

  Запрет на использование неавторизованных публичных ИИ для обработки конфиденциальных данных. Внедрение систем мониторинга для выявления несанкционированного использования ИИ-инструментов. Проведение регулярных кампаний по информированию сотрудников о рисках Shadow AI.

## Обучение и информирование сотрудников

- [ ] Обязательные тренинги по ИИ-безопасности

  Все сотрудники, использующие ИИ-инструменты в работе, обязаны проходить ежегодные тренинги по политике безопасности ИИ, этическим нормам и правилам работы с конфиденциальными данными. Тренинги будут включать практические примеры и сценарии использования.

- [ ] Каналы для сообщений о нарушениях и уязвимостях

  Сотрудники обязаны немедленно сообщать о любых потенциальных нарушениях данной политики, подозрительной активности ИИ-систем или выявленных уязвимостях в отдел информационной безопасности или своему непосредственному руководителю. Для этого предусмотрен канал связи [Указать канал, например, email: security@company.com или внутренний портал].

## Ответственность за нарушения

Несоблюдение данной политики может привести к следующим последствиям:

* Дисциплинарные меры: В зависимости от тяжести нарушения, могут быть применены дисциплинарные взыскания, вплоть до увольнения.

* Юридическая ответственность: В случаях, предусмотренных законодательством, нарушение политики может повлечь за собой гражданскую, административную или уголовную ответственность.

* Финансовые потери: Компания оставляет за собой право взыскать ущерб, причиненный в результате несоблюдения политики, в соответствии с действующим законодательством.

Каждое нарушение будет рассматриваться индивидуально, с учетом всех обстоятельств и степени причиненного ущерба.

## Приложение 1: Пример формы согласия на использование ИИ

Я, [ФИО Сотрудника], подтверждаю, что ознакомлен(а) с «Политикой безопасности при использовании ИИ в компании [Название Вашей Компании]» и обязуюсь строго соблюдать все ее положения. Я понимаю свою ответственность за безопасное и этичное использование ИИ-инструментов, а также за защиту конфиденциальных данных компании.

Дата: [Дата]

Подпись: _____

## Приложение 2: Чек-лист для оценки безопасности нового ИИ-инструмента

- [ ] Проверена ли политика конфиденциальности ИИ-сервиса?

  Соответствует ли она требованиям компании по обработке данных?
- [ ] Какие данные будут передаваться в ИИ-систему?

  Содержит ли эта информация коммерческую тайну, персональные данные или другую конфиденциальную информацию?
- [ ] Есть ли возможность контролировать и удалять введенные данные?

  Как осуществляется управление жизненным циклом данных?
- [ ] Предоставляет ли ИИ-сервис API-ключи или другие механизмы доступа?

  Как будет осуществляться контроль доступа и защита от компрометации?
- [ ] Существуют ли риски предвзятости или неэтичного поведения ИИ?

  Как будут минимизироваться эти риски?
- [ ] Какова степень прозрачности работы ИИ-модели?

  Возможно ли объяснить логику принятия решений?
- [ ] Какие меры кибербезопасности применяет провайдер ИИ-сервиса?

  Сертификаты, стандарты, аудиты.
- [ ] Требуется ли дополнительное обучение сотрудников для безопасного использования?

  Разработан ли план обучения?
- [ ] Как будет измеряться эффективность и безопасность внедрения?

  KPI, метрики.
- [ ] Получено ли одобрение от отдела информационной безопасности и руководства?

  Обязательный шаг перед внедрением.