памятка от StopPhish

БАЗОВЫЕ ПРАВИЛА БЕЗОПАСНОЙ РАБОТЫ С ИИ

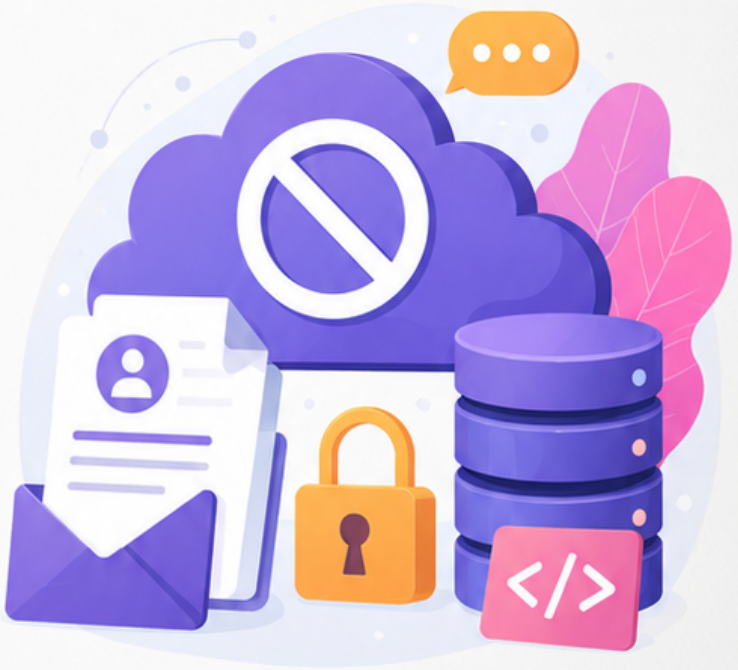
Что такое ИИ

Искусственный интеллект, или ИИ, - это программа, которая умеет обрабатывать информацию, находить закономерности и создавать ответы на основе данных, на которых её обучили.

Он может помогать с текстами, идеями, поиском ошибок, анализом и рутинными задачами. Но ИИ не понимает информацию как человек, не несет ответственности за результат и может ошибаться. Поэтому с ним нельзя обращаться как с экспертом или безопасным хранилищем данных. Всё, что вы отправляете в публичный ИИ-сервис, потенциально покидает организацию.

Главное правило: не загружайте в ИИ то, что не готовы опубликовать в открытом интернете.

Что нельзя отправлять в ИИ



Никогда не вставляйте в публичные ИИ-сервисы:

- персональные данные клиентов, коллег и партнеров;
- внутренние документы, регламенты и переписку;
- финансовые показатели до их публикации в открытом доступе;
- код программ и техническую документацию;
- коммерческую или государственную тайну;
- протоколы совещаний и рабочие заметки.

Даже если задача кажется безобидной - «сократи текст», «сделай протокол», «найди ошибки» - данные всё равно передаются стороннему сервису.

КАК ПОНЯТЬ, МОЖНО ЛИ ИСПОЛЬЗОВАТЬ ИИ

Перед отправкой информации задайте себе три вопроса:

1 Это публичная информация?

Например, статья, пресс-релиз, текст с сайта. Обычно можно публиковать.

2 Это внутренняя информация компании?

Например, черновик плана, рабочий документ, регламент. Используйте только одобренные организацией ИИ-сервисы для таких документов.

3 Это конфиденциальные данные?

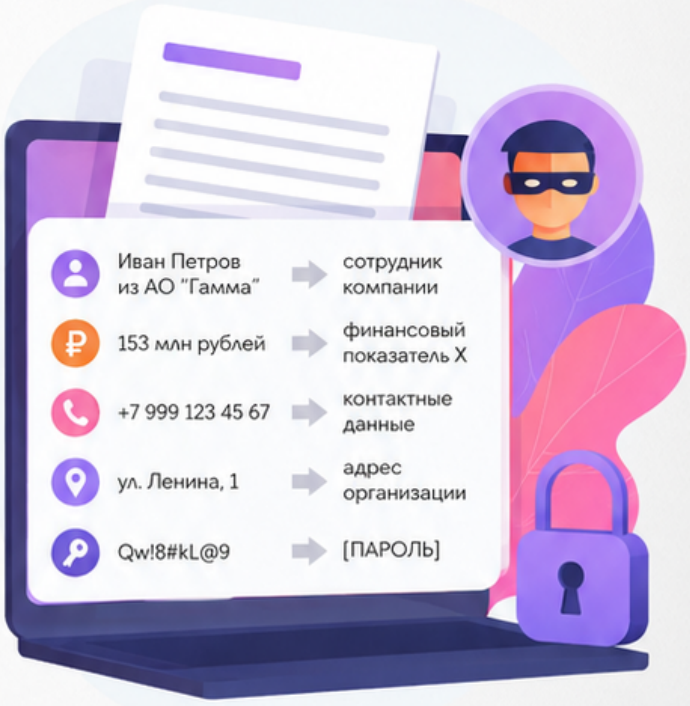
Например, пароли, персональные данные, коммерческая тайна. Отправлять в ИИ нельзя.

КАК ОБЕЗЛИЧИВАТЬ ДАННЫЕ

Если использовать публичный ИИ разрешено, уберите всё, по чему можно определить человека, организацию или проект.

Примеры замены:

- «Иван Петров из АО “Гамма”» → «сотрудник компании»;
- «выручка 153 млн рублей» → «финансовый показатель X»;
- реальные телефоны и почты → «контактные данные»;
- адреса → «адрес организации»;
- пароли и ключи → «[ПАРОЛЬ]», «[КЛЮЧ]».



! Но обезличивание не всегда спасает. Если в одном чате вы уже писали, где работаете, а потом загрузили «анонимный» отчет, данные можно сопоставить.

памятка от StopPhish

ПОЧЕМУ НЕЛЬЗЯ СЛЕПО ДОВЕРЯТЬ ИИ

ИИ может уверенно отвечать неправдой. Он способен придумать закон, источник, цитату, расчет, судебное решение или какой-то факт.

Ответственность за итоговый текст, решение или расчет несет человек, а не ИИ.

ЧТО ДЕЛАТЬ:

- проверяйте факты в первоисточниках;
- открывайте ссылки, которые дает ИИ, и проверяйте информацию;
- не используйте ИИ как единственный источник;
- просите ИИ не додумывать данные;
- задавайте конкретные вопросы и давайте для ИИ больше данных, так он ответит точнее.

Хорошая формулировка, когда вы задаете вопрос для ИИ:

> Если данных не хватает, не придумывай ответ. Задай уточняющие вопросы.

памятка от StopPhish

ЧТО ДЕЛАТЬ, ЕСЛИ ВЫ СЛУЧАЙНО ОТПРАВИЛИ ЛИШНЕЕ

Не скрывайте инцидент. Чем быстрее вы сообщите о проблеме, тем выше шанс снизить ущерб.

1. Сделайте скриншот запроса и ответа ИИ.
2. Не удаляйте чат самостоятельно.
3. Сообщите ответственному за информационную безопасность.
4. Укажите, какой сервис использовался.
5. Опишите, какие данные были переданы.
6. Если отправили пароль или доступ – срочно инициируйте замену.

5 ПРАВИЛ БЕЗОПАСНОЙ РАБОТЫ С ИИ

1. Не отправляйте в публичный ИИ конфиденциальные данные.
2. Используйте только одобренные компанией сервисы.
3. Обезличивайте информацию перед загрузкой.
4. Проверяйте факты, ссылки и расчеты.
5. Сразу сообщайте в ИБ, если отправили лишнее.

С пожеланиями безопасности, StopPhish