НЦПТИ

ИИ и конфиденциальная информация:

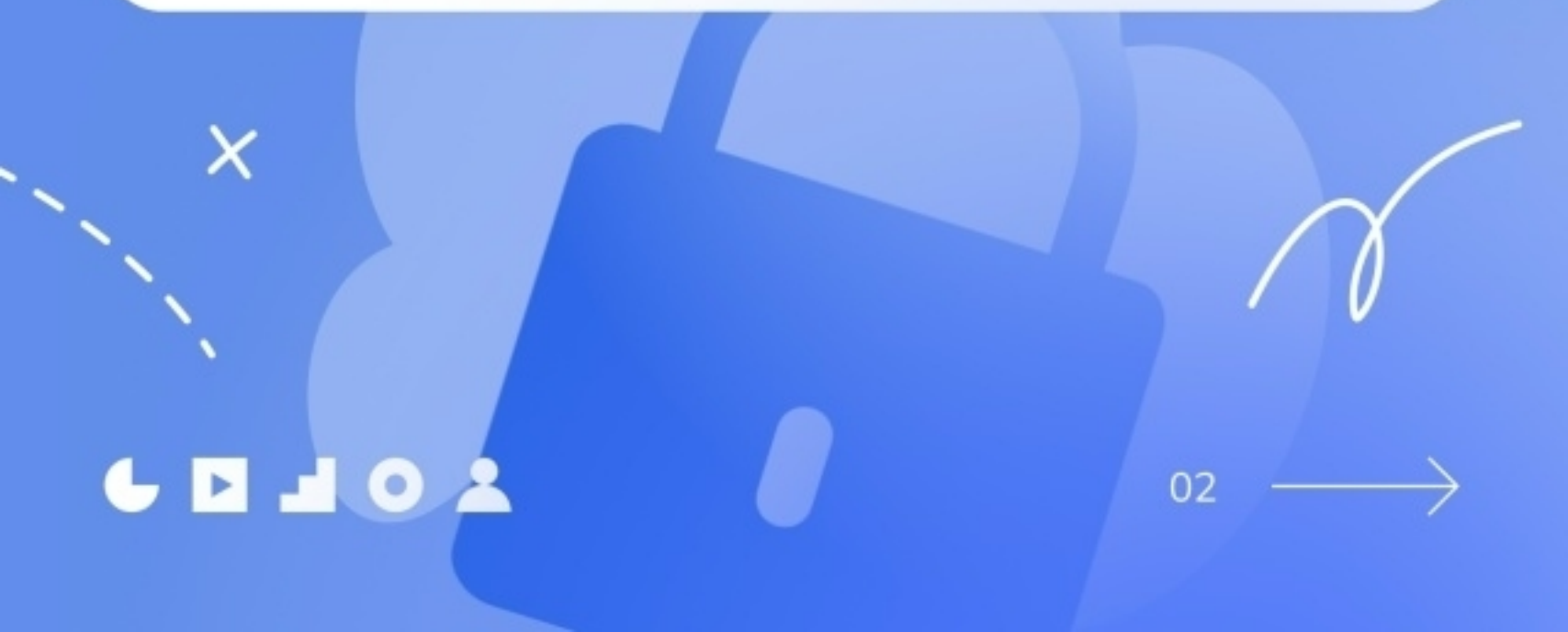
как защитить свои данные при работе с нейросетями



Конфиденциальные данные: что нужно защищать в первую очередь?

Конфиденциальная информация — это сведения, разглашение которых может нанести ущерб человеку или предприятию. К ним относятся:

- персональные данные (паспортные, контактные, медицинские, финансовые данные);
- корпоративные данные (коммерческая тайна, интеллектуальная собственность, финансовая информация, данные о клиентах и сотрудниках).



Личная информация может попадать к нейросетям следующими способами:

Через машинное обучение. Открытые ИИ-модели (например, DeepSeek) развиваются за счет обработки больших массивов информации с веб-сайтов, соцсетей, статей и документов из открытых платформ. Данные, которые пользователь присылает нейросети, также могут использоваться для обучения.

Пользователи вносят данные самостоятельно. Например, когда загружают документы для анализа информации или проверки грамотности; просят написать отчет из конкретных финансовых показателей; сообщают голосовым помощникам свое имя и адрес для вызова такси; передают номера платежных карт цифровым финансовым помощникам разных банков.

Как нейросети используют данные?

Данные, которые пользователи отправляют нейросетям, передаются разработчику сервиса и хранятся на серверах компании. Это необходимо для трех основных целей:

- обучение и улучшение модели;
- персонализация ответов под конкретного пользователя;
- обеспечение безопасности (например, выявление попыток мошенничества).

Существуют связанные с этим риски →

Утечки из-за уязвимостей. Даже крупные компании периодически сталкиваются с утечками данных. Если злоумышленники получат доступ к серверам, в их руки могут попасть паспорта, финансовые документы компаний или коммерческая тайна.

Продажа данных третьим лицам. Некоторые бесплатные ИИ-сервисы могут анализировать запросы для таргетированной рекламы и продавать обезличенные (но потенциально деанонимизируемые) данные маркетинговым компаниям.

Незаконное использование персональных данных. Злоумышленники могут применять утекшую информацию для отправки поддельных писем от имени банков/коллег, взлома аккаунтов через предоставленные пароли или шантажа, если в чат загружались компрометирующие фото или документы.

Примеры:

20 марта 2023 некоторые пользователи зарубежной нейросети смогли увидеть сообщения других людей, общавшихся с чат-ботом. Скомпрометированной оказалась и личная информация некоторых платных подписчиков сервиса: имена и фамилии, почты, платежные адреса, последние четыре цифры кредитных карт и сроки их действия.

Из-за ограничения работы зарубежных ИИ-сервисов мошенники создают фейковые сайты, предлагающие «обходной доступ». Вместо полноценных нейросетей пользователи получают урезанные версии продукта и рискуют потерять персональные и корпоративные данные. Преступники также могут использовать фишинговые схемы для кражи паролей и платежных данных клиентов.

Как обезопасить свои данные при работе с нейросетью?

1. **Изучить политику конфиденциальности сервиса** — есть ли в ней пункт об использовании данных (в том числе для дообучения) и можно ли отменить эту опцию. Например, в настройках аккаунта Яндекс можно ограничить передачу запросов для обучения ИИ.
2. **Обезличивать данные,** которые будут переданы, не указывать чувствительную информацию, которая может причинить вред вам или вашей компании при ее утечке.
3. **Выработать в корпоративной среде для своей команды правила работы с внешними сервисами:** что можно загружать, а что нельзя.
4. Пользоваться наиболее популярными российскими ИИ-сервисами вместо новых и неизвестных.
5. **Проверять ссылки ИИ-сервисов** — под видом известной платформы может скрываться фейковый сайт, ворующий пароли и платежные данные.