

Практическое руководство
для женщин и девушек

Искусственный интеллект

дипфейки и защита вашего образа



Введение

Независимо от того, являешься ли ты женщиной в политике или в деловой среде, инфлюенсеркой, преподавательницей, врачом, студенткой или работаешь в любой другой сфере, ты можешь столкнуться с тем, что станешь объектом атак с использованием фото-, видео- или аудиоматериалов, созданных с помощью инструментов искусственного интеллекта.

Дипфейк – так называется фотография, видеоролик или аудиозапись, в которых используются данные о тебе (твое изображение и/или голос), но которые не отражают реальность. Иными словами, тебя там не было, и ты не делала и не говорила того, что представлено в этом материале.

Использование сгенерированного контента с целью манипуляции, шантажа, запугивания или подрыва репутации человека является ЦИФРОВЫМ НАСИЛИЕМ.



Да, ты не обязана кому-либо когда-либо доказывать, что фотография или видео с тобой на самом деле являются дипфейком. Тем не менее важно знать, как защитить свой образ в случае, если дипфейк, направленный против тебя, распространяется в публичном онлайн-пространстве.

Будь **ИНФОРМИРОВАННОЙ** и **ПОДГОТОВЛЕННОЙ**, чтобы уметь отличать реальный контент от контента, сгенерированного с помощью искусственного интеллекта.

Ниже приведены несколько признаков, которые могут помочь прояснить ситуацию в случае дипфейка, созданного с целью ввести людей в заблуждение:

Фотографии

- ✓ *Неправильные пропорции тела;*
- ✓ *Большее или меньшее количество пальцев;*
- ✓ *Стертые или размытые детали;*
- ✓ *Несоответствия фона;*
- ✓ *Нереалистичные или чрезмерно идеализированные*

Видеоизображения

- ✓ Пустой, отсутствующий взгляд;
- ✓ Скванные, неестественные движения;
- ✓ Ненатуральная мимика;
- ✓ Слишком ровный цвет кожи без естественных несовершенств;
- ✓ Нечёткие детали в области глаз, носа или рта;
- ✓ Ощущение наложения изображений в моменты, когда отдельные элементы становятся размытыми при движении рук, рта и т.п.;
- ✓ Тени или освещение, которые выглядят неестественно либо не должны присутствовать.

Звук

- ✓ Качество звука отличается, даже если речь идёт об одном и том же контексте; могут появляться резкие перепады.
- ✓ Можно заметить, что на протяжении видео или в отдельных фрагментах движение губ не синхронизируется с произносимыми словами.
- ✓ Голос может звучать неестественно, не соответствовать человеку на записи или иметь иную интонацию, чем предполагает контекст слов.

В продолжение представляем три факта о дипфейках и твоей безопасности в цифровом пространстве.

1. Даже если в настоящее время с помощью внимательного взгляда можно распознать сгенерированный контент, искусственный интеллект развивается настолько быстро, что в ближайшем будущем мелкие детали, которые сегодня помогают отличать подделку, могут исчезнуть.

Когда ты публикуешь или распространяешь контент с собой, старайся не использовать версии самого высокого качества. Таким образом повышается вероятность того, что эти материалы не будут использованы для создания контента, воспроизводящего твой образ, поведение или голос.

2. Дипфейки используются не только с целью манипулирования общественным мнением, например в случае женщины, занимающей высокую должность. Они могут быть очень легко использованы и для манипуляции или запугивания женщины, не имеющей значимого публичного статуса, которая находилась в абьюзивных отношениях, когда бывший партнёр решает «отомстить» тем или иным способом.

Веди учёт контента, который касается тебя и хранится на устройствах других людей, даже если на данный момент эти люди тебе очень близки. Договаривайся или проси удалить такие материалы во время отношений или после их завершения, если существует возможность общения с соответствующим человеком.

3. Существует множество инструментов и приложений, доступных обычным пользователям (без специальных технических навыков), которые позволяют создавать контент, воспроизводящий реальность. Технология работает вместо человека, нуждаясь лишь в данных о тебе. Чем больше информации можно извлечь из фотографий и видеоматериалов, тем ближе дипфейк будет к твоему реальному образу и поведению.

Проверь, какой контент, связанный с тобой, доступен в онлайн-среде, и проанализируй контексты, в которых ты представлена. Ограничь доступ к этим материалам, удалив их или установив ограниченные настройки доступа в социальных сетях и на других платформах.

Чего не следует делать

если сгенерированный контент используется, чтобы **ЗАПУГАТЬ, МАНИПУЛИРОВАТЬ, ШАНТАЖИРОВАТЬ, УГРОЖАТЬ** или **ПРИНУЖДАТЬ ТЕБЯ ДЕЛАТЬ ТО, С ЧЕМ ТЫ НЕ СОГЛАСНА:**

Не удаляй доказательства!

Сообщения, фотографии, видео, комментарии – всё, что подтверждает существование или распространение данного дипфейка, необходимо задокументировать и сохранить с помощью скриншотов или записей.

Не выполняй требования, навязанные агрессором!

Как правило, агрессоры не прекращают свои действия и не соблюдают обещания остановиться даже после выполнения выдвинутых ими условий. Кроме того, они не удаляют контент, связанный с тобой, или сгенерированные материалы, воспроизводящие твой образ. Их тенденция – использовать такие материалы как инструменты причинения вреда всякий раз, когда это им выгодно.

Поэтому важно твёрдо сказать **НЕТ** и проанализировать дальнейшие шаги для своей защиты как в онлайн-, так и в офлайн-среде.

Не сомневайся – найдутся люди, готовые тебе помочь!

Когда речь идёт о твоей безопасности, страх перед тем, что подумают другие, может только навредить. Обратись с уверенностью к человеку, который выслушает тебя с терпением и пониманием. Обсудите вместе, какими будут следующие шаги, и попроси поддержки, чтобы сообщить о произошедшем

Не нужно радикально менять своё поведение в онлайн-среде!

Ты имеешь право присутствовать в интернете! Ты имеешь право продвигать свои услуги в онлайн-среде. Ты имеешь право делиться, публиковать и рассказывать о своей жизни.

Агрессоры, которые создали и/или распространили контент с тобой без твоего согласия, не должны становиться причиной ограничения твоего онлайн-присутствия. Если ты действительно этого хочешь, продолжай быть активной онлайн, но делай это

- ✓ **Проанализируй возможные риски:** кражу личности, онлайн-преследование, манипуляцию, мошенничество, захват контроля над личными аккаунтами и т. д.
- ✓ **Настрой параметры конфиденциальности!** Ограничь доступ подозрительных лиц к своей информации. Удаляй, ограничивай или блокируй аккаунты, которые вызывают у тебя недоверие или доставляют дискомфорт.
- ✓ **Сообщай о фейковых аккаунтах,** аккаунтах присвоивших чужую идентичность, контенте, который пропагандирует насилие или дискриминацию, разжигает ненависть или поощряет вредоносное поведение – то есть об аккаунтах и материалах, нарушающих стандарты онлайн-сообществ.
- ✓ **Сообщай об изображениях,** содержащих сексуальное насилие над детьми, на платформе siguronline.md.
- ✓ **Сообщай о распространении интимных изображений** взрослых (18+) без согласия на платформе StopNCII.org. Прочитай здесь подробнее о том, как можно предотвратить распространение твоих интимных изображений, а также их распространение без согласия.

СВЯЖИТЕСЬ С НАМИ!

- ☎ 0 8008 8008 - Телефон Доверия для женщин и девушек
- ✉ trustline@lastrada.md
- 💬 www.eviolenta.md онлайн-чат

eviolenta.md

Данный материал был разработан Международным Центром «La Strada» и Фондом ООН в области народонаселения (UNFPA в Молдове) при финансовой поддержке Великобритании.

