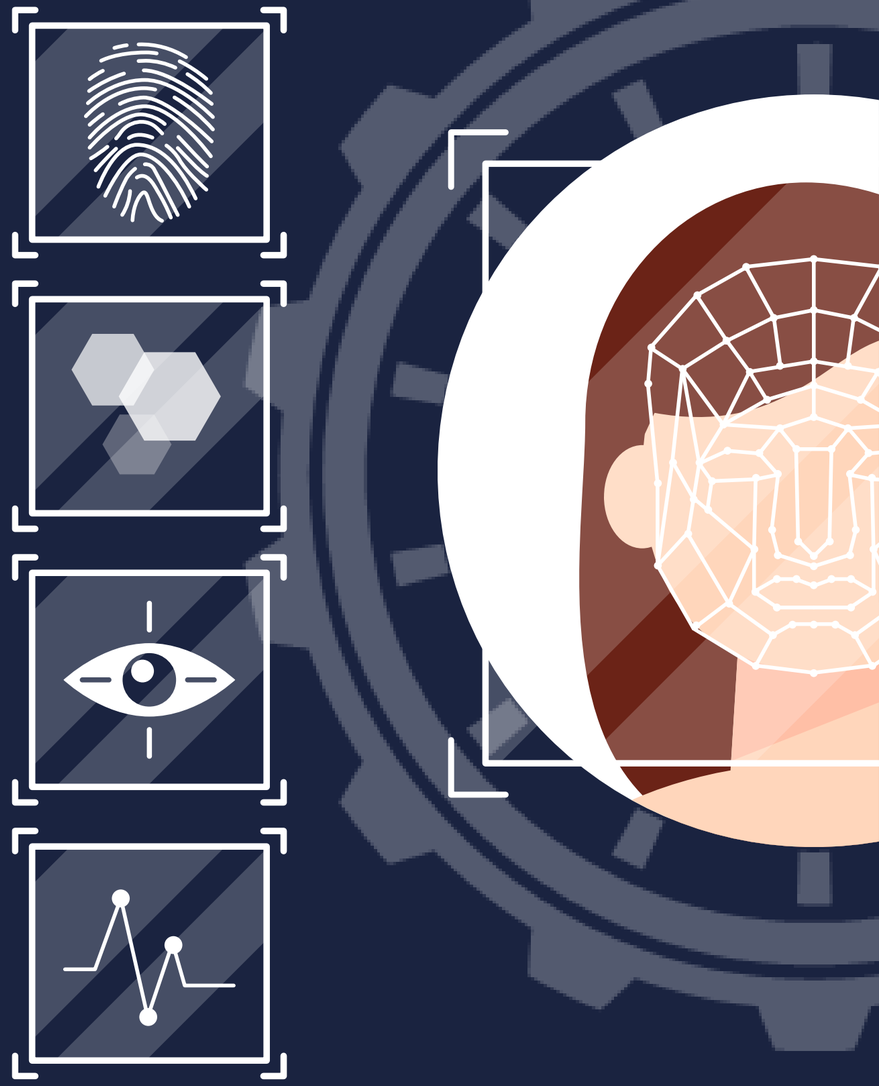
Поймай меня, если сможешь! Дипфейки в биометрии: угроза или искусство?



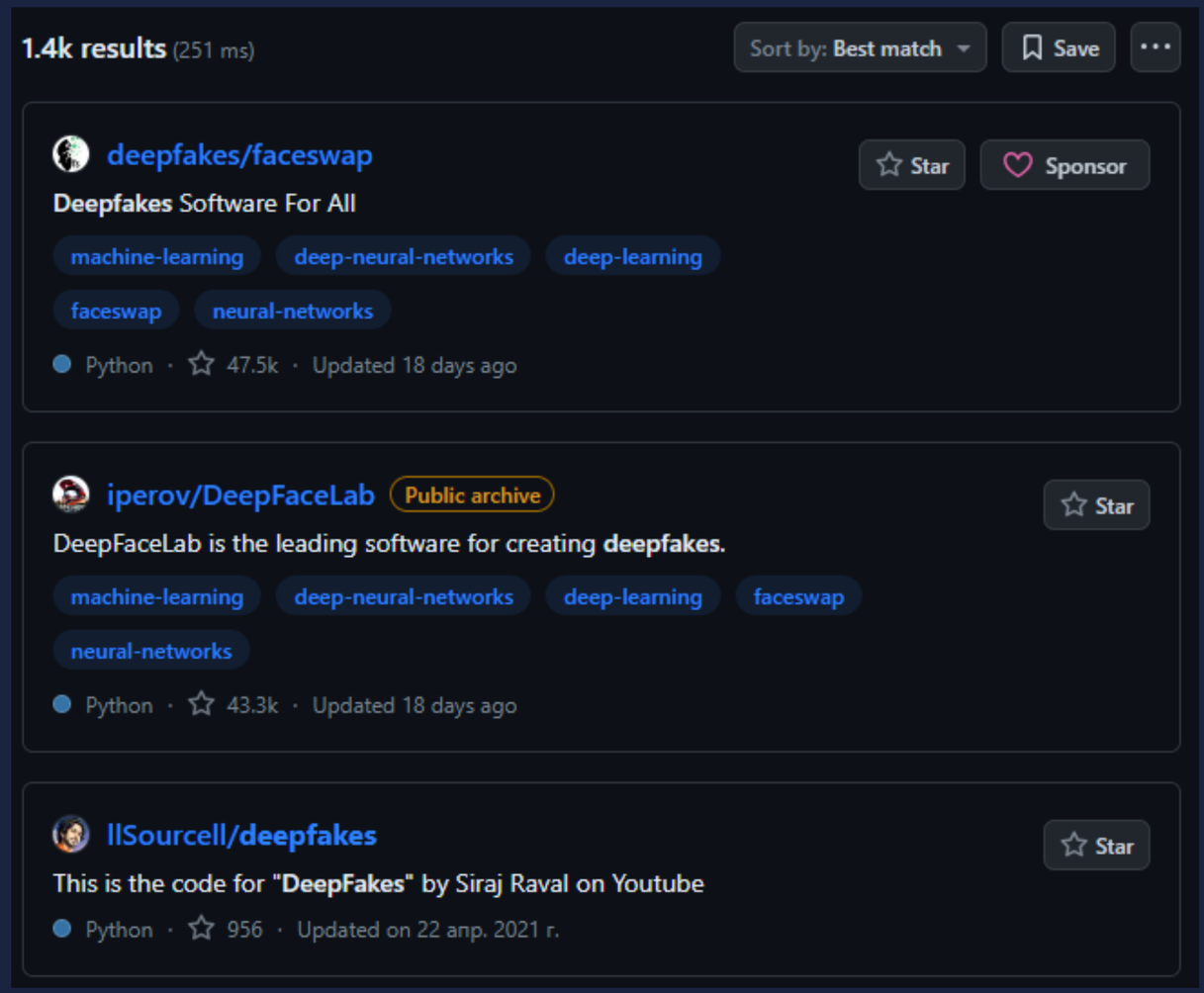
Почему дипфейки набирают популярность?

РОСТ БИОМЕТРИЧЕСКИХ ТЕХНОЛОГИЙ

По мере того, как биометрия находит все более широкое применение, мошенники становятся более изобретательными в методах атаки на биометрию. Количество случаев мошенничества с использованием биометрических данных увеличивается ежегодно в 2 раза.

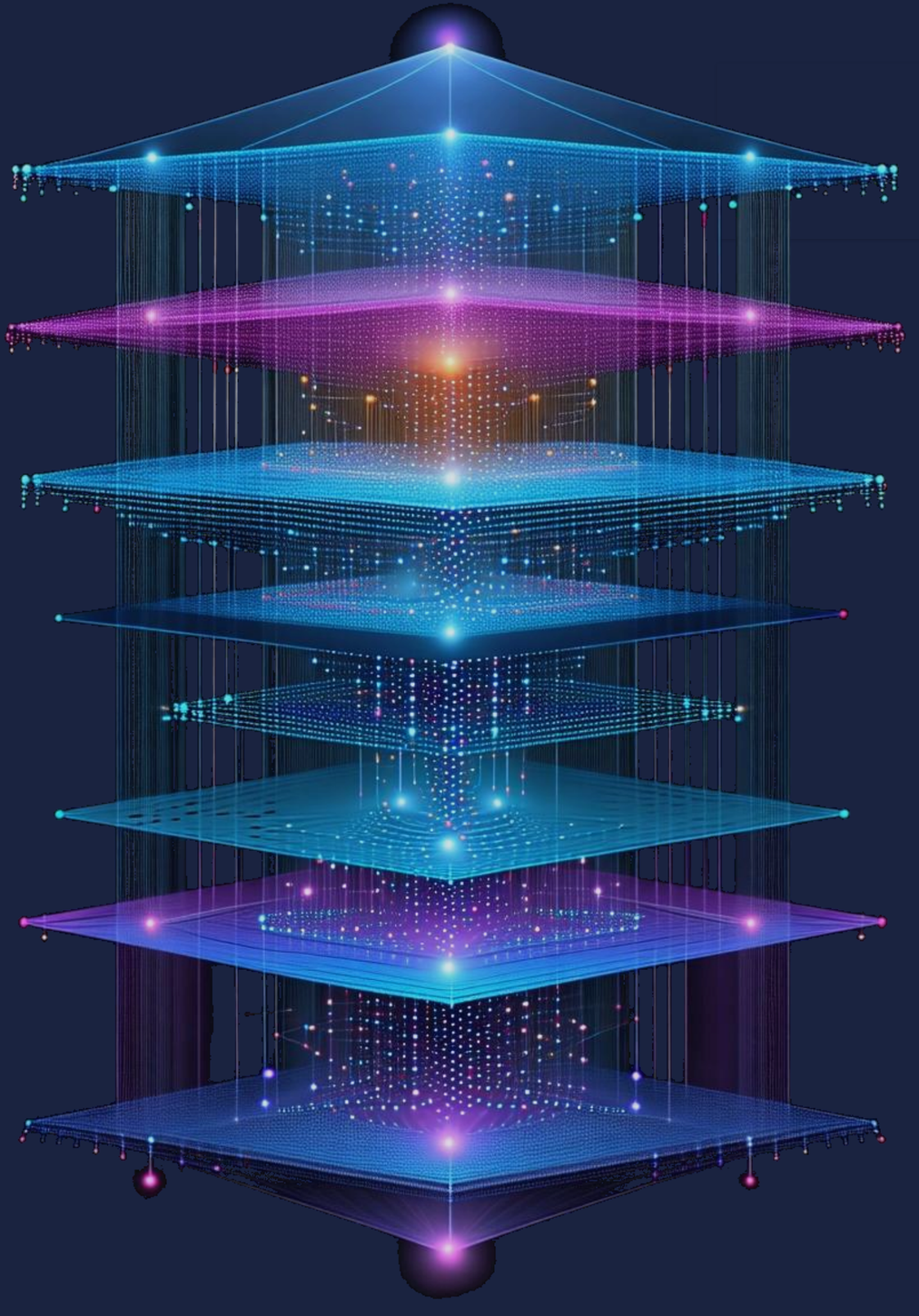
ИСПОЛЬЗОВАНИЕ НЕЙРОСЕТЕЙ

Доступность алгоритмов, аппаратных ресурсов и баз данных упрощает реализацию атак. Реалистичность изображений, создаваемых генеративно-состязательными сетями (GAN) растет, а время создания дипфейков и требования к железу снижаются.



Нейронные сети в биометрии

Нейронные сети обучаются на данных.
Для эффективного обучения нейронных сетей требуются большие наборы данных, содержащие как легитимные, так и атакующие примеры.
Правильная настройка гиперпараметров необходима для оптимизации производительности и минимизации ложных срабатываний при обнаружении атак.



Атаки на биометрические системы

пассивная попытка «самозванца»

злоумышленник не прилагает никаких усилий, чтобы быть распознанным как другой человек, а просто использует свои собственные биометрические характеристики

активная попытка «самозванца»

Злоумышленник пытается совпасть с сохраненным шаблоном другого человека, предъявляя поддельный биометрический образец, либо намеренно изменяя свои собственные биометрические характеристики

Атаки на биометрические системы

пассивная попытка «самозванца»

злоумышленник не прилагает никаких усилий, чтобы быть распознанным как другой человек, а просто использует свои собственные биометрические характеристики

biometric recognition

активная попытка «самозванца»

Злоумышленник пытается совпасть с сохраненным шаблоном другого человека, предъявляя поддельный биометрический образец, либо намеренно изменяя свои собственные биометрические характеристики

liveness detection
deepfake detection

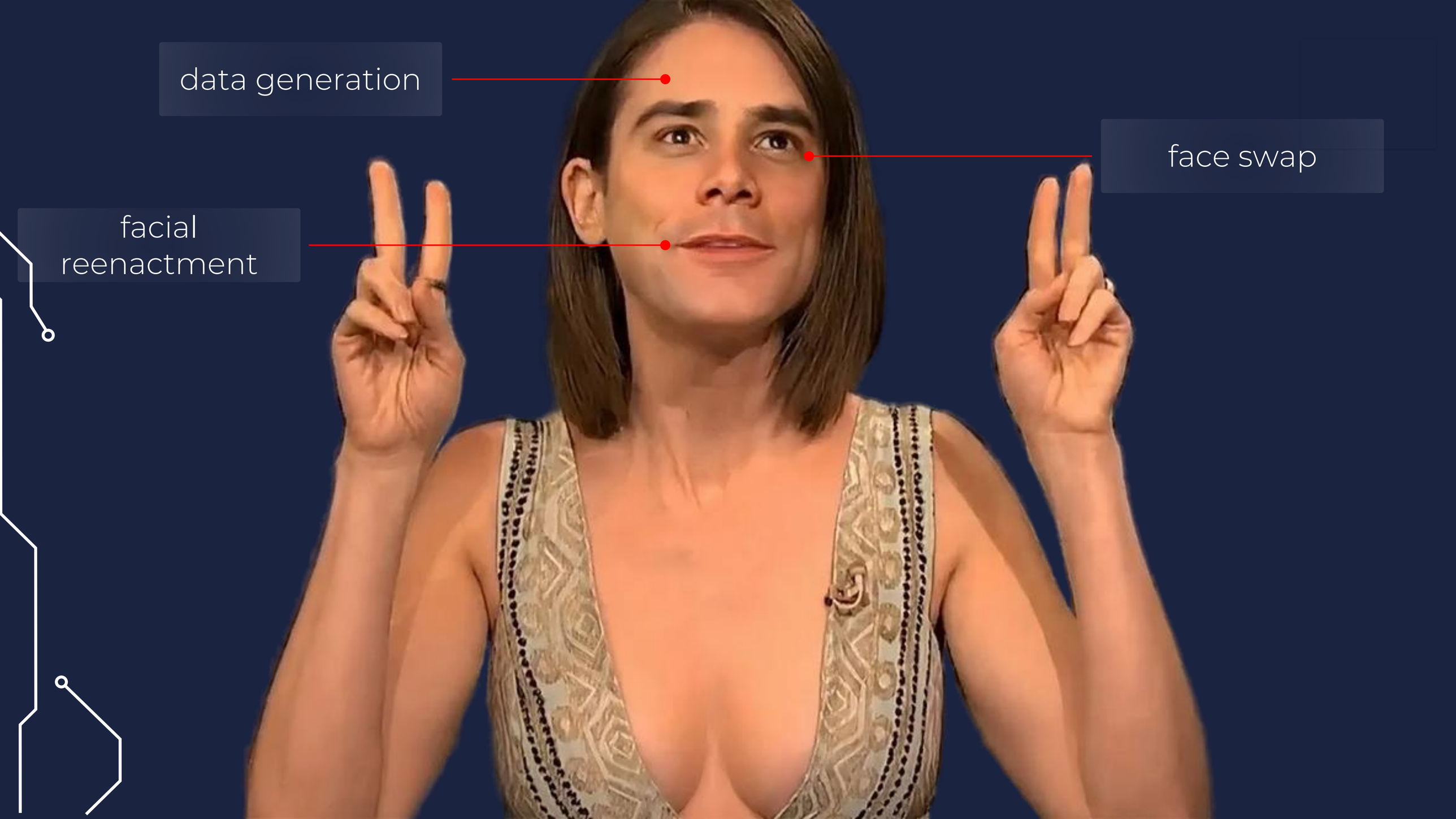
Liveness detection vs. Deepfake detection

Спуфинг предполагает использование поддельных биометрических признаков для получения несанкционированного доступа. Например, использование распечатанных фотографий или силиконовых масок для обмана биометрических систем.

Дипфейки— это синтетические данные, созданные с использованием технологий искусственного интеллекта (ИИ) и машинного обучения, которые реалистично заменяют внешность, голос, действия или иные характеристики физического лица или создают их с нуля.

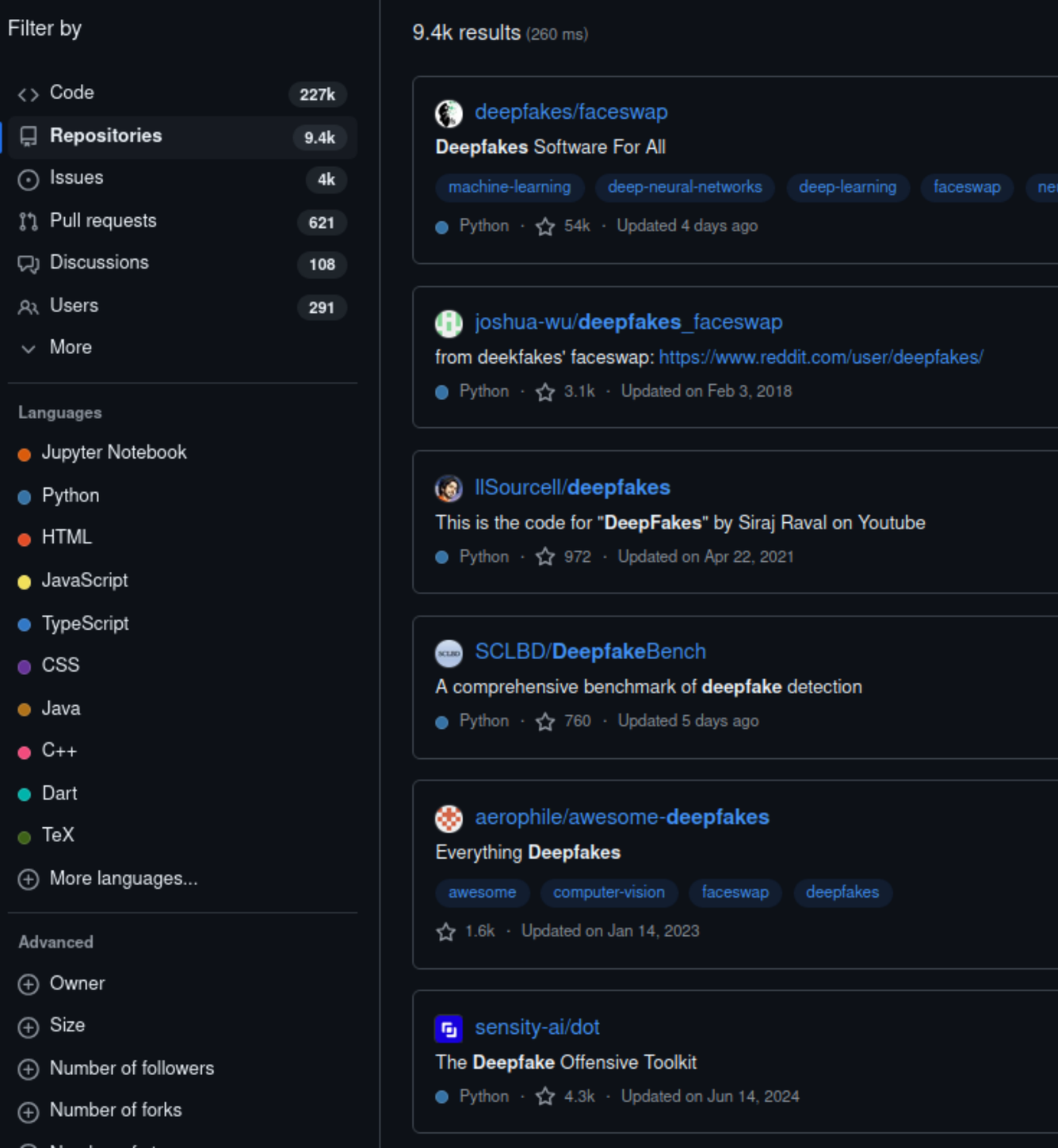
Что такое дипфейк?

синтетические данные, созданные с использованием технологий искусственного интеллекта (ИИ) и машинного обучения, которые реалистично заменяют внешность, голос, действия или иные характеристики физического лица или создают их с нуля


data generation
face swap
facial reenactment

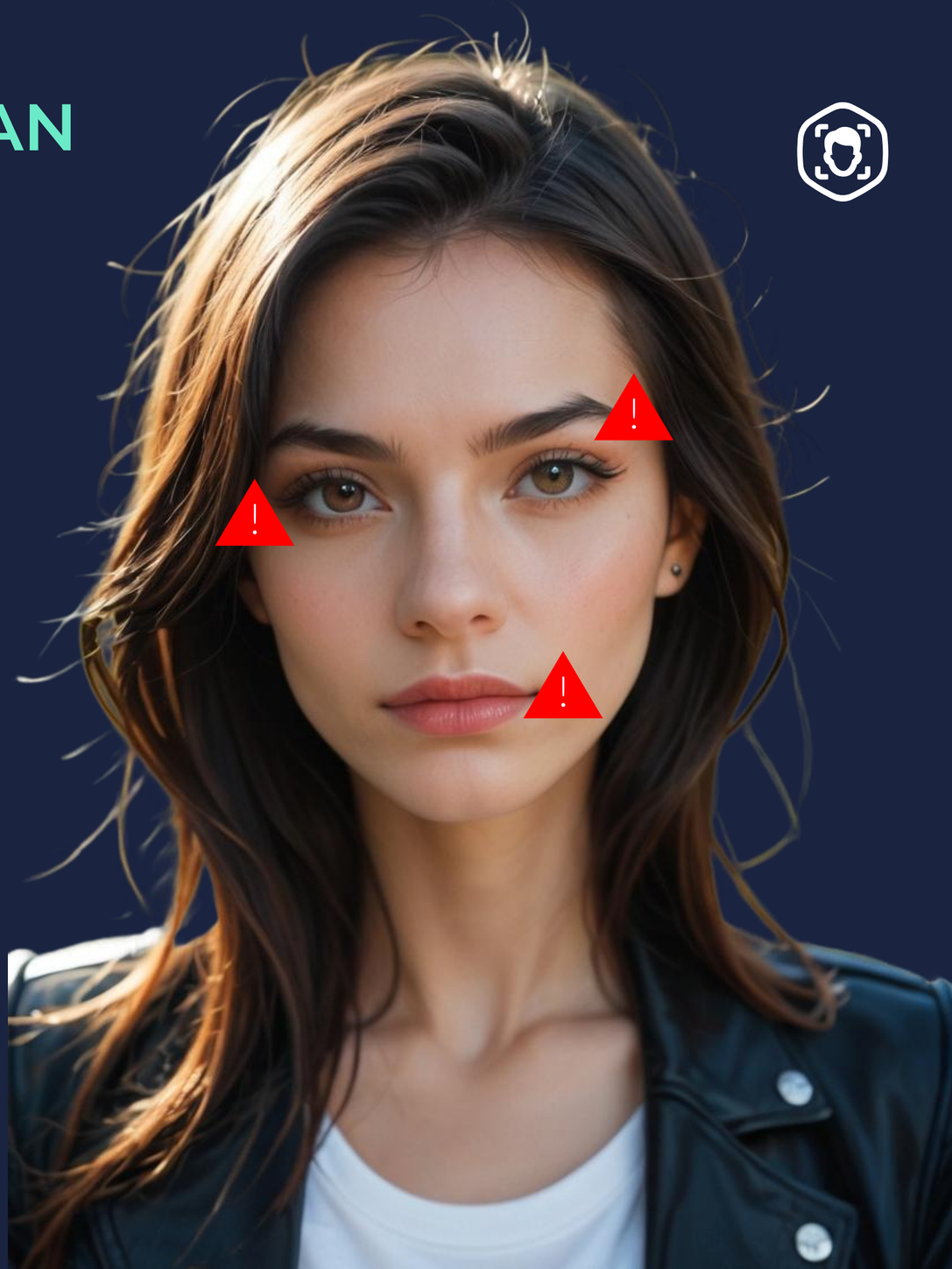
Как создать дипфейк?

Генеративные нейронные сети стали неотъемлемой частью создания контента, изменив нашу повседневную жизнь, рекламу, киноиндустрию, политику и т.д.



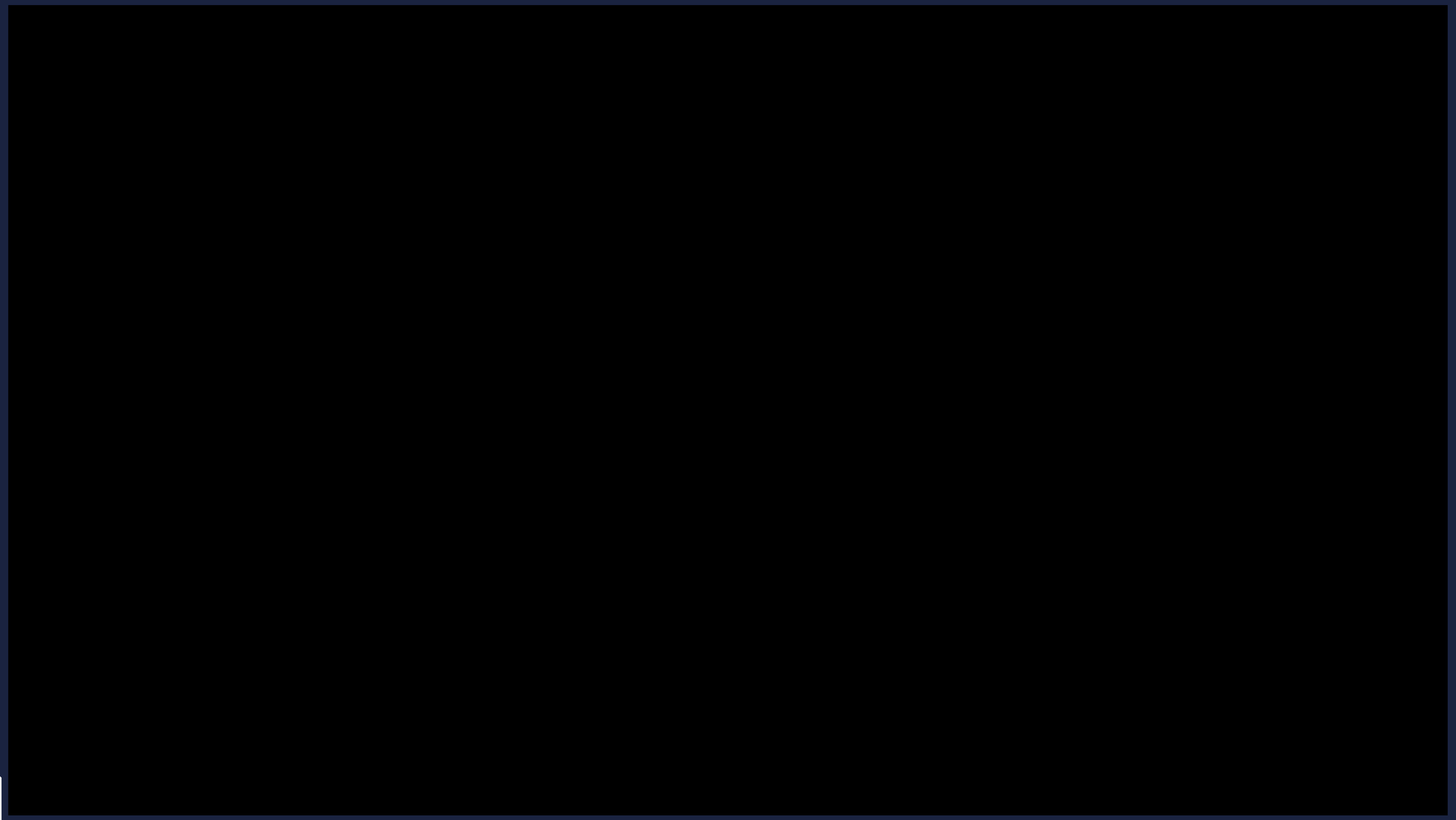
Как детектировать дипфейк: GAN против дипфейк детекторов

Нейронные сети обучаются на данных. Для эффективного обучения нейронных сетей для детекции дипфейков требуются большие наборы данных, включающие как реальные образцы, так и дипфейки. Для создания реалистичного дипфейка также требуются как можно больше данных



Как детектировать дипфейк?

Как детектировать дипфейк?



Сериал «ПМЖейсон», Agenda Media Group

Почему детектировать дипфейк непросто?

Эволюция генеративных сетей

для генерации дипфейков

Малое количество реальных примеров

дипфейков, использованных для реальных атак

Увеличение доступности аппаратного обеспечения

для обучения генеративных нейронных сетей

Почему детектировать дипфейк непросто?

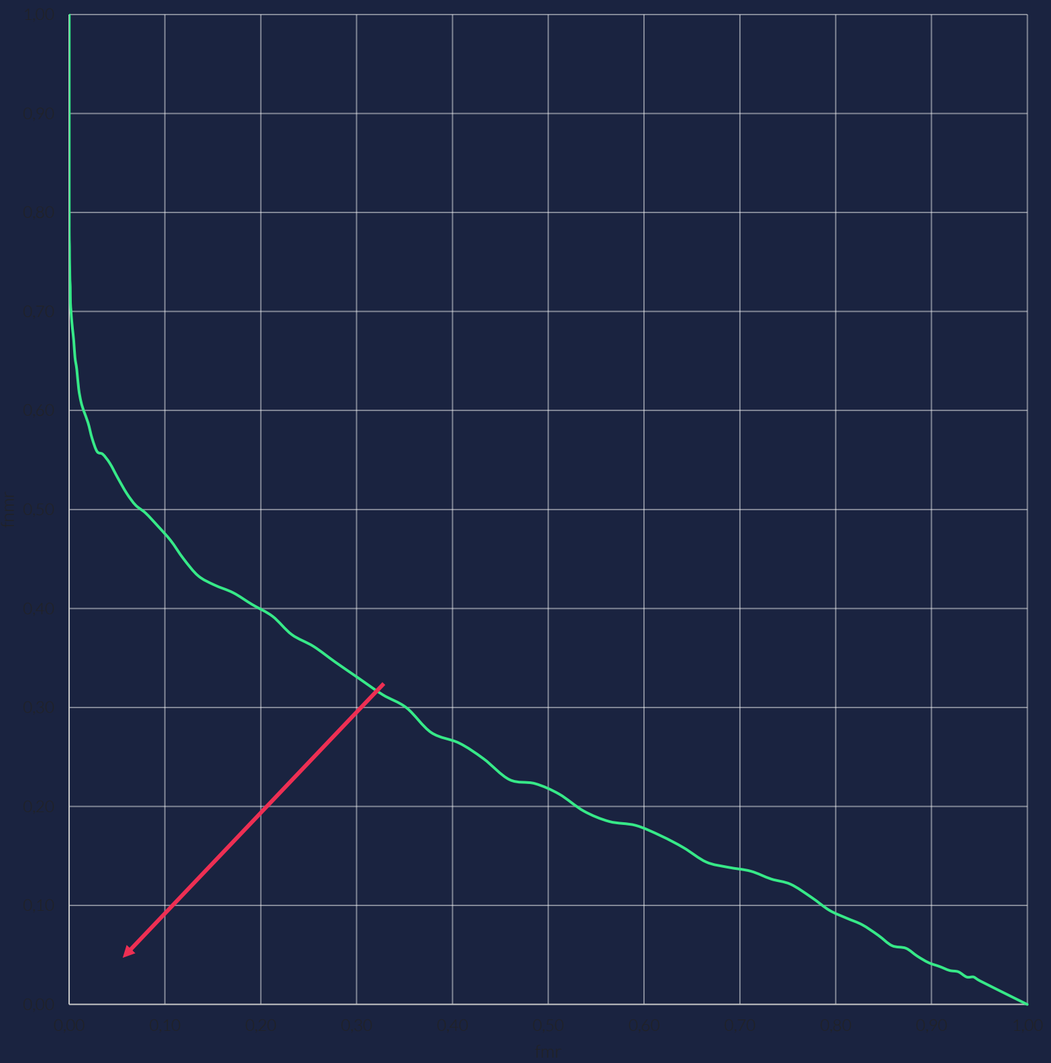
Одним из методов борьбы с дипфейками является маркировка искусственного контента. Но такие данные не подходят для обучения дипфейк детекторов, т.к. могут привести к отравлению обучающей базы данных

ОЦЕНКА АЛГОРИТМОВ

FAR
Доля попыток атак, которые будут ошибочно приняты

FRR
Доля попыток подлинного лица, которые будут ошибочно отвергнуты

Изменяя пороговое значение, мы получаем пары значений ошибок



Мы на связи!

Официальный портал
ebs.ru

Социальные сети «Биометрия РФ»

Бессонова Наталья

natalya.bessonova@ebs.ru