Meta* создала систему защиты от голосовых клонов.

Компания Meta* разработала новую технологию под названием AudioSeal , позволяющую встраивать скрытые сигналы, известные как водяные знаки, в аудиозаписи, сгенерированные с помощью искусственного интеллекта. Эта система призвана помочь в обнаружении ИИ-созданного контента в Интернете и противодействии растущей проблеме дезинформации и мошенничества с использованием инструментов для клонирования голоса.

Хади Эльсахар, научный сотрудник Мета, утверждает, что AudioSeal – первый инструмент, способный точно определять, какие части аудиозаписи, например, в подкасте, могли быть созданы с помощью ИИ.

Однако существует ряд серьёзных оговорок. Мета пока не планирует применять водяные знаки к аудио, созданному с помощью их инструментов ИИ. Водяные знаки для аудио ещё не получили широкого распространения, и не существует единого согласованного отраслевого стандарта. Кроме того, водяные знаки для ИИ-сгенерированного контента обычно легко подделать или удалить.

Быстрое обнаружение и возможность точно определять, какие элементы аудиофайла созданы ИИ, имеют решающее значение для полезности системы, говорит Эльсахар. Его команда добилась от 90% до 100% точности в распознавании водяных знаков, что значительно лучше предыдущих попыток.

AudioSeal доступен бесплатно на GitHub . Любой может скачать его и использовать для добавления водяных знаков в ИИ-сгенерированные аудиоклипы. В будущем он может быть интегрирован с моделями генерации ИИ-речи, чтобы автоматически применяться к любой созданной с их помощью речи.

AudioSeal создана с использованием двух нейронных сетей. Первая генерирует специальные сигналы водяных знаков, которые можно встраивать в аудиодорожки. Эти сигналы полностью неразличимы для человеческого уха, но могут быть быстро обнаружены второй нейросетью. В отличие от существующих методов, где необходимо проверять аудио по секундным фрагментам на наличие водяного знака, что крайне медленно и трудоемко, AudioSeal работает совершенно иначе. Система встраивает водяной знак по всей длине аудиодорожки, распределяя его равномерно. Благодаря этому водяной знак становится «локализованным» и его можно обнаружить даже в том случае, если аудиофайл был обрезан, отредактирован или из него были удалены фрагменты.

Но существуют дополнительные серьёзные недостатки, которые необходимо преодолеть, прежде чем такие аудиоводяные знаки можно будет принять повсеместно. Исследователи Meta обнаружили, что чем больше информации раскрывается об алгоритме водяного знака, тем более уязвимым он становится. Кроме того, система требует, чтобы люди добровольно добавляли водяной знак к своим аудиофайлам.

Несмотря на популярность решений с водяными знаками в технологическом секторе, некоторые эксперты скептически относятся к тому, что они действительно повысят общественное доверие к информации из-за их уязвимости к злонамеренному удалению и подделке.

*Компания Meta и её продукты признаны экстремистскими, их деятельность запрещена на территории РФ.